

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日 Date of Application: 2003年 8月 6日

出願番号 Application Number: 特願 2003-206168

ST. 10/C] : [JP 2003-206168]

願人 Applicant(s): 株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

USSN 10/712,031
MATTINGLY, STANGER, MALUR & BRUNDIDGE
(703) 684-1120
TSM-34

2003年10月15日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



BEST AVAILABLE COPY

出証番号 出証特 2003-30845

【書類名】 特許願

【整理番号】 K03009841A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 12/00

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

【氏名】 岩見 直子

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

【氏名】 山本 康友

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

【氏名】 藤本 和久

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社 日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 ストレージ装置

【特許請求の範囲】

【請求項 1】

計算機と接続されるインターフェース部と、
ファイル操作の処理を行う第一の制御部と
記憶装置に対するデータの読み書きの処理を行う第二の制御部と、
前記インターフェース部、前記第一の制御部及び前記第二の制御部とを相互に
接続する接続網とを有し、
前記インターフェース部は、前記計算機から送信されたフレームの転送先を前
記第一の制御部及び前記第二の制御部のうちいずれか一つから選択し、前記接続
網を介して前記フレームを選択した制御部へ転送することを特徴とするストレー
ジ装置。

【請求項 2】

前記第一の制御部は、前記インターフェース部から前記フレームを受信して前
記フレームで指定されたファイル操作を行う場合、前記第二の制御部を介して前
記記憶装置に対するデータの読み書きを実行することを特徴とする請求項 1 記載
のストレージ装置。

【請求項 3】

他のストレージ装置と接続する第二のインターフェース部を有し、
前記インターフェース部は、前記フレームの転送先を前記第一の制御部、前記
第二の制御部又は前記第二のインターフェース部から選択することを特徴とする
請求項 2 記載のストレージ装置。

【請求項 4】

前記第一の制御部は、前記インターフェース部から前記フレームを受信して前
記フレームで指定されたファイル操作を行う場合、前記第二のインターフェース
部を介して前記他のストレージ装置に対するデータの読み書きを実行することを
特徴とする請求項 3 記載のストレージ装置。

【請求項 5】

前記第一の制御部を複数有し、

前記インターフェース部は、前記フレームがファイル操作を要求するコマンドを含むフレームであった場合、前記複数の第一の制御部から所定の第一の制御部を選択し、選択された前記第一の制御部へ前記フレームを前記接続網を介して送信することを特徴とする請求項1記載のストレージ装置。

【請求項6】

前記インターフェース部は、前記複数の第一の制御部と前記計算機から受信するフレームに含まれる識別子との対応関係に関する情報を保持し、前記フレームを受信した際に前記情報に基づいて前記フレームを転送する先の第一の制御部を決定することを特徴とする請求項5記載のストレージ装置。

【請求項7】

前記フレームを受信した前記第一の制御部の指示により、前記インターフェース部及び前記第二の制御部は、前記接続網を介して前記第一の制御部の処理に関するデータを送受信することを特徴とする請求項1記載のストレージ装置。

【請求項8】

前記フレームを受信した前記第一の制御部の指示により、前記インターフェース部及び前記第二のインターフェース部は、前記接続網を介して前記第一の制御部の処理に関するデータを送受信することを特徴とする請求項4記載のストレージ装置。

【請求項9】

前記情報には、前記フレームを受信する前記インターフェース部が有する一つのポートに複数の前記第一の制御部が対応することを示す情報が含まれていることを特徴とする請求項6記載のストレージ装置。

【請求項10】

前記インターフェース部を複数有し、前記複数のインターフェース部で受信されたフレームが前記第一の制御部に転送されることを特徴とする請求項1記載のストレージ装置。

【請求項11】

更に管理部を有し、前記インターフェース部は、前記管理部の指示に従って、

前記情報の内容を変更し、前記フレームの転送先を変更することを特徴とする請求項 9 記載のストレージ装置。

【請求項 12】

前記インターフェース部を複数有し、

前記管理部の指示に従って、前記インターフェース部で行っていた処理を他の前記インターフェース部に引継ぐ手段を有することを特徴とする請求項 11 記載のストレージ装置。

【請求項 13】

前記管理部の指示は、前記インターフェース部の障害を該ストレージ装置が有する各装置が検出した際に行われ、前記管理部は、前記インターフェース部の障害の際に処理を引継ぐ前記他のインターフェース部の情報を有することを特徴とする請求項 12 記載のストレージ装置。

【請求項 14】

前記第一の制御部を複数有し、

前記管理部の指示に従って、前記第一の制御部で行っていた処理を他の前記第一の制御部に引継ぐ手段を有することを特徴とする請求項 12 記載のストレージ装置。

【請求項 15】

前記管理部の指示は、前記第一の制御部の障害を該ストレージ装置が有する各装置が検出した際に行われ、前記管理部は、前記第一の制御部の障害の際に処理を引継ぐ前記他の第一の制御部の情報を有することを特徴とする請求項 14 記載のストレージ装置。

【請求項 16】

前記第二の制御部は、前記第二のインターフェース部を介して前記他のストレージ装置を制御することを特徴とする請求項 4 記載のストレージ装置。

【請求項 17】

前記第二の制御部は、キャッシュメモリ及びディスク装置を有することを特徴とする請求項 1 から 16 に記載のうちいずれか一つのストレージ装置。

【請求項 18】

計算機と接続されるインターフェース部と、

ファイル操作の処理を行う制御部と

記憶装置に対するデータの読み書きの処理を行う第二のインターフェース部と

前記インターフェース部、前記第二のインターフェース部及び前記制御部とを相互に接続する接続網とを有し、

前記インターフェース部は、前記計算機から送信されたフレームの転送先を前記第一の制御部及び前記第二のインターフェース部のうちいずれか一つから選択し、前記接続網を介して前記フレームを選択した前記第一の制御部又は前記第二のインターフェース部へ転送することを特徴とするストレージ装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、計算機システムにおいて計算機によって使用されるデータを格納するストレージ装置、特にネットワーク経由でファイルアクセスされるストレージ装置に関する。

【0002】

【従来の技術】

近年、計算機で取り扱われるデータ量は飛躍的に増大し、これに応じてデータを格納するストレージ装置の大容量化が進んでいる。大規模かつ低コストな大型ストレージを実現する方法として、特許文献1に開示されているクラスタ構成のストレージシステムがある。

【0003】

クラスタ構成ストレージシステムは、比較的小構成のストレージ（以下「ストレージノード」）複数台をスイッチ等で相互に結合した構成を有する。このような構成では、ストレージノードを必要とされる記憶容量に合わせて複数結合してクラスタ構成ストレージシステムを構成することができ、小容量から大容量まで様々な容量を持つストレージシステムをスケーラブルに実現できる。

【0004】

一方、ストレージ装置の種類の一つとして、ネットワークアタッチドストレージ（以下「NAS」）がある。これは、ネットワーク経由でのファイルアクセスを受け付けることが可能なストレージ装置で、実際には、ファイルシステムを提供する計算機（以下「NASサーバ」）及びディスク装置等の記憶装置から構成される。尚、「ファイル」とは、ファイルシステム等でユーザが認識するデータの一塊を指す。

NASは、使用者がファイルアクセスを直接利用できるため使いやすいストレージ装置である。しかし、NASへアクセスするためのプロトコル（NFS等）が、TCP/IP（Transmission Control Protocol/Internet Protocol）の使用を前提としているため、NASへのアクセス頻度が上がるにつれ、NASにおけるTCP/IPの処理負荷が問題となってきた。この問題を解決する技術として、TCP/IPの処理をNASサーバから分離して単独のハードウェア（以下「TCP/IP処理部」）とし、TCP/IP処理部とNASサーバをスイッチで接続する技術が特許文献2に開示されている。特許文献2では、計算機が発行したコマンドは、TCP/IP処理部でTCP/IPレイヤまで処理され、その後NASサーバ部で処理される。NASサーバが記憶装置をアクセスする場合は、Fiber Channel Switching Module経由でSANに接続された記憶装置を利用する。

【0005】

【特許文献1】

米国特許6,256,740号

【特許文献2】

PCT国際公開 02/069166号公報

【0006】

【発明が解決しようとする課題】

特許文献2に開示されている技術では、アクセス対象となるNASサーバとNASサーバに付与されたネットワークアドレスが設定されるTCP/IP処理部のポートとの関係は1:1で一意に決定される。このため、NASサーバに対するアクセス頻度が上がったり、送受するデータ量が増加すると、TCP/IP処理部又はNASサーバの処理性能がネックとなり、レスポンスが低下する。また、TCP/IP処理部又はNAS

サーバに障害が発生した場合、障害が発生したNASの使用を継続できない。

又、上記の問題は、単純に特許文献1のクラスタ構成ストレージシステムにNASを導入するだけでは解決しない。

【0007】

本発明の目的は、NASの使用負荷が上がった場合にも、レスポンスが低下しないNASを提供することにある。

また、本発明の別の目的は、NASに障害が発生した場合にも障害が発生したNASへのアクセスを継続できるNASを提供することにある。

【0008】

【課題を解決するための手段】

本発明の一実施形態はストレージ装置であり、以下の構成を有する。ストレージ装置は、計算機との間でやり取りされるフレーム（コマンド、データ等含む。パケット等でも良い）の通信プロトコルを処理するインターフェース部（以下「I/F処理部」）と、ファイル操作（ファイルへのデータの書き込み、読み出し、作成、削除等）の処理を行う第一の制御部（以下「NAS部」）と、記憶装置を制御する第二の制御部（以下「キャッシュ制御部」）と、各部を相互に接続する接続網とを有する。本構成において、計算機が送信したフレームを受信したI/F処理部は、受信したフレームをその内容に基づいてNAS処理部に送信したり、キャッシュ制御部に転送したりする。その際I/F処理部は、I/F処理部に割り当てられたアドレスとNAS部との対応関係の情報を用いる。

【0009】

又、他の実施形態として、上述のストレージ装置が別のストレージ装置と接続される第二のI/F処理部を有する構成としても良い。尚、この場合、更に、上述の構成からキャッシュ制御部を省略した構成も考えられる。この場合、フレームは、第一のI/F処理部及び第二のI/F処理部を介して他のストレージ装置へ転送される。

【0010】

一方、フレームを受け取ったNAS部は、状態に従い、キャッシュ制御部または、第二のI/F処理部経由で別のストレージ装置にアクセスして所定の処理を行つ

た後、フレームに対する応答をI/F処理部に送信する。その際NAS部は、ファイルと当該ファイルが記憶されているキャッシュ制御部又は他のストレージ装置の対応関係の情報を用いる。

【0011】

なお、別の実施形態として、I/F処理部とNAS部を複数備える構成とし、I/F処理部とNAS部との対応関係を適宜設定することで、処理負荷の分散を図ったり、障害が発生した時に処理を引き継ぐ構成とすることが考えられる。

また、別の実施形態として、キャッシュ制御部がキャッシュメモリ及びディスク装置を有さない構成としても良い。

【0012】

また、本実施例の好ましい実施形態として、I/F処理部に割り当てられたアドレスとNAS部との対応関係、NAS部が処理するファイルと当該ファイルが記憶されているキャッシュ制御部もしくは他のストレージ装置の対応関係等、各種情報を管理する管理用の計算機（以下、「管理サーバ」）又は同様の管理部をストレージ装置に設ける（又はストレージ装置に接続する）構成がある。管理サーバが、I/F処理部に割り当てられたアドレスとNAS部との対応関係の変更命令をストレージ装置に出すことで、処理負荷の分散を図ったり、障害が発生した時に処理を引き継ぐよう構成変更をすることが可能な構成とする。

【0013】

【発明の実施の形態】

図1は、本発明を適用したシステムの一実施形態の構成を示す図である。システムは、一台以上の計算機（以下「ホスト」）100、一台以上のスイッチ120、ストレージ装置（以下「ストレージ」）130、ストレージ130とスイッチ120経由で接続される一台以上のストレージ（以下「外部ストレージ」）180、管理サーバ110及び管理端末190とを有する。

【0014】

ホスト100、スイッチ120、ストレージ130及び外部ストレージ180は、IPネットワーク175を介して管理サーバ110に接続され、管理サーバ110で動作する図示しない管理ソフトウェアによって統合管理される。なお、

本実施形態では、ストレージ130は管理端末190を介して管理サーバ110に接続する形態をとるものとするが、ストレージ130が直接IPネットワークに接続される構成であっても良い。

【0015】

ホスト100は、CPU101やメモリ102などを有する計算機であり、ディスク装置や光磁気ディスク装置などの記憶装置103に格納されたオペレーティングシステムやアプリケーションプログラムなどのソフトウェアをメモリ102に読み上げ、CPU101がメモリ102から読み出して実行することで、所定の処理を行う。又、ホスト100は、キーボードやマウスなどの入力装置104やディスプレイ等の出力装置105を具備し、入出力装置104でホスト管理者などからの入力を受け付け、出力装置105にCPUから指示された情報が出力される。

【0016】

さらにホスト100は、ストレージ130等外部の装置とネットワーク176を介してデータを送受するための一以上のポート107と、管理サーバとIPネットワーク175を介してデータ送受するための一以上のインターフェース制御部106を有する。

【0017】

管理サーバ110は、ホスト100と同様の計算機であり、システム全体の運用・保守管理といった所定の処理を行う。CPU111によって管理ソフトウェアが実行されると、管理サーバ110はインターフェース制御部116からIPネットワーク175を介して、システム内の各機器から構成情報、リソース利用率、性能監視情報などを収集する。そして、管理サーバ110は収集したそれらの情報を出力装置115に出力してストレージの管理者等に提示する。また、管理サーバは、入力装置114を介してストレージ管理者からの指示を受信し、受信した運用・保守指示をインターフェース制御部116を介して各機器に送信する。

【0018】

スイッチ120は、複数のポート121及びインターフェース制御部を有する。ポート121は、ホスト100のポート107とIPネットワーク176を介し

て、または、ストレージ130のポート141とIPネットワーク177を介して接続される。又、インターフェース制御部123は、IPネットワーク175を介して管理サーバ110に接続される。スイッチ120は、IPネットワーク内に存在するスイッチが有する経路制御等一般的な処理を行う。

【0019】

ストレージ130はクラスタ構成を有する。具体的には、ストレージ130は複数のプロトコル変換部140、複数のキャッシュ制御部150、1つ以上のNAS部145及び構成管理部160を有し、これらは相互結合網170で相互に接続されている。

【0020】

プロトコル変換部140は、複数のポート141、1つ以上の制御プロセッサ142、メモリ143、及び相互結合網170と接続される転送制御部144を有する。プロトコル変換部140は、ポート141から受信したコマンドを解析し、アクセス対象となるデバイスやNAS部145を特定し、転送制御部144より相互結合網170を介して適当なキャッシュ変換部150やNAS部145等へコマンドやデータを転送する。尚、以下「デバイス」とは、ストレージが提供する物理的又は論理的な記憶領域のことと言う。

【0021】

また、プロトコル変換部140は、ポート141を介して外部ストレージ180など別のストレージと接続されることが可能である。この場合、外部ストレージ180と接続されるプロトコル変換部140は、別のプロトコル変換部140から受信する入出力要求やキャッシュ制御部150から受信する入出力要求を外部ストレージ180へ送信する。これにより、ホスト100又はNAS部145は、外部ストレージ180にデータを書き込んだり（以下「ライト」）、外部ストレージ180からデータを読み出したり（以下「リード」）することができる。

【0022】

なお本実施形態では、ポート141はIPネットワークで主に使用されるTCP/IPと、SCSIを上位プロトコルとしたFiber Channel（以下「FC」）に対応しているとする。NASへのアクセスに使用されるNFS等の通信プロトコルは、TCP/IPプロト

コルの上位プロトコルとして位置する。本実施形態では、記述の簡単のため、NASへのアクセスに使用されるファイルアクセスプロトコルをNFSとするが、CIFS等別のプロトコルでもかまわないし、複数のファイルアクセスプロトコルをサポートしてもかまわない。また、IPネットワークを介してストレージアクセス（デバイスのアクセス）を行う場合に使用されるiSCSI(Internet SCSI)等も、TCP/IPプロトコルの上位プロトコルとして位置し、プロトコル変換部140は、NFS等NASのアクセスに使用されるプロトコルに加え、iSCSI等ストレージが有するデバイスをアクセスするためのプロトコルに対応する。

【0023】

ホスト100がストレージ130又は外部ストレージ180のデバイスをプロックデバイスとして利用したい場合、FC又はIPのネットワーク経由でFCやiSCSI等のプロトコルを用いてストレージ130と通信を行う。

なお、ポート141がFCのプロトコルに対応しない構成でも良い。また、NFS等NASのアクセスに使用されるプロトコルのみに対応する構成でも良い。

【0024】

キャッシング制御部150は、1つ以上のポート156、各々ポートに接続される1台以上のディスク装置157、1つ以上の制御プロセッサ152、各々の制御プロセッサ152に対応するメモリ153、1つ以上のディスクキャッシング154、1つ以上の制御メモリ155及び相互結合網170と接続される転送制御部151を有する。

【0025】

制御プロセッサ152は、相互結合網170を介して転送制御部151で受信した、同じキャッシング制御部150内に存在するディスク装置157への入出力要求を処理する。また制御プロセッサ152は、キャッシング制御部150がホスト100またはNAS部145に対してディスクアレイのように複数のディスク装置157を1つ又は複数の下位論理デバイスとして提供している場合に、下位論理デバイスとディスク装置157との間の対応関係の管理や、下位論理デバイスに対するアクセス要求をディスク装置157へのアクセス要求に変換する処理などを行う。更に制御プロセッサ152はデータ複製やデータ再配置などの各種処

理を実行する。

【0026】

ディスクキャッシュ154には、ホスト100またはNAS部145からのアクセス要求に対する処理速度を高めるため、ディスク装置157から頻繁に読み出されるデータが予め格納されたり、ホスト100から受信したライトデータが一時的に格納される。尚、ディスクキャッシュ154を用いたライトアフタ、即ちホスト100またはNAS部145から受信したライト用のデータ（以下「ライトデータ」）がディスクキャッシュ154に格納された後、ディスク装置157に実際に書き込まれる前にホスト100またはNAS部145に対しライト要求に対する応答を返す場合、ディスクキャッシュ154に格納されているライトデータがディスク装置157に書き込まれる前に消失することを防止するため、ディスクキャッシュ154をバッテリバックアップなどにより不揮発メモリとしたり、媒体障害への耐性向上のため二重化するなど、ディスクキャッシュ154の可用性を向上させておくことが望ましい。

【0027】

制御メモリ155には、キャッシング制御部150が、ディスク装置157、複数のディスク装置157を組み合わせて構成される物理デバイス又はストレージ130にプロトコル変換部140を介して接続される外部ストレージ180内のデバイス（以下「外部デバイス」）の管理や、外部デバイス又は物理デバイスと下位論理デバイスの対応関係の管理を行うために必要な情報が格納される。なお、制御メモリ155に格納されている制御情報が消失すると、ホスト100やNAS部145はキャッシング制御部150が有するディスク装置157に格納されているデータへアクセスできなくなるため、制御メモリ155はバッテリバックアップなどにより不揮発メモリとしたり、媒体障害への耐性向上のため二重化するなど、高可用化のための構成を有することが望ましい。

【0028】

また、キャッシング制御部150の別の構成として、ポート156とディスク157を有さず、外部ストレージ180のデバイスのみを制御する構成としても良い。

【0029】

NAS部145は、一または複数のCPU147、各々の制御プロセッサに対応する一または複数のメモリ149、記憶装置148及び相互結合網170と接続される転送制御部146を有する。記憶装置148に格納された制御プログラムをメモリ149に読み上げ、これをCPU147が実行することで、NAS部145は所定の処理を行う。

【0030】

CPU147は、相互結合網170を介して転送制御部146より受信したコマンドを処理し、コマンドを送信してきたプロトコル変換部140へ応答コマンドを送信する。また、CPU147は、コマンド処理時にデバイスアクセスが必要になった場合（ファイルへのデータの書き込みや読み出し）、そのデバイスを提供（より具体的には、NAS部で実行されているファイルシステムにマウントされるデバイスを提供する）しているキャッシュ制御部150又は外部ストレージ180に対し入出力要求を転送制御部146を介して送信し、その応答を相互結合網170を介して転送制御部146より受信する。尚、デバイスへアクセスを行う場合、NAS部145は、受信したファイルアクセスのコマンドをそのファイルのデータが格納されるデバイスのブロックへのアクセスコマンドに変換する。

【0031】

構成管理部160は、一又は複数の制御プロセッサ162、各々の制御プロセッサに対応する一又は複数のメモリ163、一又は複数の制御メモリ164、記憶装置165、相互結合網170と接続される転送制御部161及びインターフェース制御部166を有する。記憶装置165に格納された制御プログラムをメモリ163に読み上げ、これを制御プロセッサ162が実行することで、ストレージ130の構成管理や障害管理のための所定の処理が実行される。

【0032】

制御プロセッサ162は、インターフェース制御部166を介して接続される管理端末190にストレージ管理者へ提示する構成情報を送信したり、管理者から管理端末190に入力された保守・運用指示を管理端末190から受信して、受領した指示に従い、ストレージ130の構成変更などを行う。ストレージ130

の構成情報は制御メモリ164に格納される。制御メモリ164に格納された構成情報はプロトコル変換部140の制御プロセッサ142、キャッシュ制御部150の制御プロセッサ152、NAS部145の制御プロセッサ147によって参照されたり更新されたりすることができる。これにより、ストレージ130内のプロトコル変換部140及びキャッシュ制御部150間で構成情報を共有することができる。

【0033】

なお、構成管理部160が障害などにより動作不可に陥った場合、ストレージ130全体へのアクセスが不可能になるため、構成管理部160内の各構成要素を二重化しておくか、もしくは構成管理部160自体をストレージ130内に複数搭載して構成管理部160自体の二重化をしておくことが望ましい。また、管理端末190から個々のキャッシュ制御部150へのI/Fを別途設け、構成管理部160が行う制御を個々のキャッシュ制御部150と管理端末190で分担することで、構成管理部160を制御メモリ164のみの構成にすることも可能である。さらに、制御メモリ164内の情報を個々のキャッシュ制御部150の制御メモリ155で保持することで、構成管理部160を省略することも可能である。

【0034】

相互結合網170は、プロトコル変換部140、キャッシュ制御部150及び構成管理部160を相互に接続し、これらの各装置間でのデータ、制御情報及び構成情報の送受信を実行する。この相互結合網170により、構成管理部160がプロトコル変換部140、キャッシュ制御部150およびNAS部145にストレージ130の構成情報を配布したり、プロトコル変換部140、キャッシュ制御部150及びNAS部145から構成情報を取得してストレージ130の構成を管理したりすることができる。

【0035】

また、相互結合網170がプロトコル変換部140、キャッシュ制御部150及びNAS部145との間のアクセス要求を転送するので、ホスト100はプロトコル変換部140の任意のポート141からNAS部145やキャッシュ制御部1

50に属する任意の上位論理デバイスへアクセスすることが可能となる。なお、可用性向上の観点から相互結合網170も多重化されていることが望ましい。

【0036】

管理端末190は、CPU192、メモリ193、記憶装置194、構成管理部160と接続するインターフェース制御部191、IPネットワーク175と接続するインターフェース制御部197、ストレージ管理者からの入力を受け付ける入力装置195及びストレージ管理者にストレージ130の構成情報や管理情報を出力するディスプレイ等の出力装置196を有する計算機である。CPU192は記憶装置194に格納されているストレージ管理プログラムをメモリ193に読み出して、これを実行することにより、構成情報の参照、構成変更の指示、特定機能の動作指示などを行い、ストレージ130の保守運用に関して、ストレージ管理者もしくは管理サーバ110とストレージ130間のインターフェースとなる。

【0037】

なお、管理端末190を省略して、ストレージ130を直接管理サーバ110へ接続し、ストレージ130を管理サーバ110で動作する管理ソフトウェアを用いて管理してもよい。

【0038】

外部ストレージ180は、一又は複数のポート181、制御プロセッサ182、メモリ183、ディスクキャッシュ184、一又は複数のディスク装置186及び各々ディスク装置186に接続される一又は複数のポート185を有する。外部ストレージ180は、各々が有するポート181及びスイッチ120を介してストレージ130に接続される。又、外部ストレージ180が有するディスク装置186から構成される論理的な記憶領域である外部デバイスは、ストレージ130によりストレージ130が有する上位論理デバイスとしてNAS部145またはホスト100へ提供される。

【0039】

なお、システムはスイッチ120の他にFCスイッチを有し、FCスイッチ経由で外部ストレージ180がストレージ130と接続されてもよい。また、スイッチを経由せず直接ストレージ130と外部ストレージ180とを接続してもよい。

また、外部ストレージ180を設けず使用しないシステム構成としてもよい。

【0040】

制御プロセッサ182はメモリ183に格納されているプログラムを実行することにより、ポート181から受信したディスク装置186への入出力要求を処理する。本実施形態では外部ストレージ180を、クラスタ構成を有さず、ストレージ130より小規模な構成のストレージ装置とするが、ストレージ130と同じ構成を有する同規模のストレージ装置であってもかまわない。また、外部ストレージ180がディスクキャッシュ184や制御プロセッサ182の無いJBOD (Just Bounce Of Disk) のような単体ディスク装置グループであっても構わない。

【0041】

以下、本実施形態のシステムの動作概要について簡単に説明する。

ストレージ130の有するディスク装置157または外部デバイスからなる上位論理デバイスの一部もしくは全ては、ファイルアクセス処理を行うNAS部145に提供される。

【0042】

ホスト100がストレージ130に記憶されるファイルを使用するためにストレージ130に送信したコマンド（ファイルアクセスのコマンド）は、ストレージ130のポート141の内、ファイルアクセス用（NFS）のアドレスが割り当てられているポート141に到着する。そのポート141を有するプロトコル変換部140は、そのコマンドを解析し、そのコマンドを処理すべきNAS部145を選択し、そのコマンドを相互結合網170経由で選択したNAS部145に送信する。

【0043】

コマンドを受信したNAS部145は、そのコマンドで指定されるファイルに対する処理を行う。コマンドの処理中にデバイス（具体的にはそのデバイスに格納されているファイル）へのアクセスが必要となった場合、そのデバイスがキャッシュ制御部150が提供する上位論理デバイスであれば、NAS部145は当該キャッシュ制御部150へアクセスコマンドを送信し、必要なデータを得たり、デ

ータをキャッシュ制御部150に格納したりする。尚、上位論理デバイスに外部デバイスが含まれる場合、キャッシュ制御部145は、プロトコル変換部145を介してコマンドを外部ストレージ180へ送信する。また、NAS部145が外部ストレージ180が提供するデバイスを直接利用している場合、NAS部145は、当該外部ストレージが接続されているプロトコル変換部140にアクセスコマンドを送信し、必要なデータを得たり、データを外部ストレージ180が提供するデバイスに格納する。

【0044】

コマンドの処理を終了したNAS部145は、応答コマンドを作成し、相互結合網170を経由して先に受信したコマンドの送信元のプロトコル変換部140に応答コマンドを送信する。応答コマンドを受信したプロトコル変換部140は、応答コマンドをホスト100へ送信する。

【0045】

なお、以降の説明では、表現の簡略化のため、プロトコル変換部140をPA、キャッシュ制御部150をCA、構成管理部160をMA、管理端末190をST、NAS部145をAAと表記する。

【0046】

また、本発明の実施形態では、ストレージ130内の各ディスク装置、物理デバイス、下位又は上位論理デバイスの関係（以下「デバイス階層」）を次の通りとする。

まず、CA内でディスク装置157複数台によるディスクアレイが構成され、このディスクアレイはCAによって物理デバイスとして管理される。また、PAに接続された外部ストレージ180が提供する外部デバイスは、PAで認識された後、MAにて管理される。

【0047】

更にCAでは、CA内に存在する物理デバイス又は当該CAを介してアクセスされる外部デバイスに対して下位論理デバイスを割り当てる。下位論理デバイスは各CA内において管理される論理デバイスであり、その番号は各CA毎に独立に管理される。下位論理デバイスは、MAによって上位論理デバイスと対応付け

られ、ストレージ130が提供する論理的な記憶装置（デバイス）としてホスト100またはAAに提供される。

【0048】

即ち、ホスト100またはAAが認識するのはストレージ130の上位論理デバイスであり、ホスト100は上位論理デバイスを識別するための情報を用いて、ストレージ130若しくは外部ストレージ180に格納されているデータにアクセスする。ただし、AAは、管理情報に従いPAを介して直接外部ストレージ180の外部デバイスにアクセスすることもできる。この場合、外部デバイスは特定のCAには割り当てられていない。

【0049】

尚、個々のディスク装置157を1つの物理デバイスおよび1つの下位論理デバイス及び上位論理デバイスとしてもよい。また、複数の物理デバイスを1つ又は複数の下位論理デバイスに対応付けたり、複数の下位論理デバイスに1つ又は複数の上位論理デバイスを対応付けても良い。

以下、本システムの動作を説明する前に、本システムの各構成要素が有する情報について説明する。

【0050】

図2は、ストレージ130の各構成要素の制御メモリおよびメモリに格納される、制御情報、管理情報とストレージ制御処理のためのプログラムの一例を示したソフトウェア構成図である。

【0051】

ストレージ130のデバイス階層を管理するための管理情報として、CAの制御メモリ155に下位論理デバイス管理情報201、物理デバイス管理情報202及びキャッシュ管理情報203が、MAの制御メモリ164に上位論理デバイス管理情報204、外部デバイス管理情報205及びLUパス管理情報206が格納されている。

【0052】

CA内の制御メモリ155およびMAに格納されている各制御情報は、各CA、各PA、各AA、及びMA内の各制御プロセッサから参照・更新可能であるが

、その際に相互接続網 170などを介したアクセスが必要となる。よって、処理性能向上のため、各制御プロセッサで実行される処理に必要な制御情報の複製が、各部位（即ち、CA、PA、AA及びMA）内のメモリに保持される。

【0053】

各部位は、当該部位が管理する制御情報が構成変更により更新された場合は、相互接続網 170を介してその旨を他の部位に通知し、最新情報を当該部位の制御メモリから他の部位のメモリへ取り込ませる。なお、制御情報が更新された場合に更新を他の部位に知らせる別の方法として、例えば、各部位は、自部位内の制御メモリに保持した構成情報毎に制御メモリ上に更新有無を示すフラグを設け、各部位の制御プロセッサは処理開始時もしくは各構成情報を参照する毎に同フラグを参照して更新有無をチェックする方法などが考えられる。

【0054】

なお、各部位のメモリには上述の制御情報の複製に加えて、各部位内の制御プロセッサで動作する制御プログラムが格納される。

【0055】

NASの処理（ファイルアクセス処理、障害時のフェイルオーバ等）にかかるる管理情報としては、MAの制御メモリ 164にNAS管理情報 223、ポート管理情報 224が格納されている。

【0056】

PAのメモリ 143には、上位論理デバイス管理情報 204の全て又は当該PAの処理にかかるる情報の複製 214、外部デバイス管理情報 205の全て又は当該PAの処理にかかるる情報の複製 215、LUパス管理情報 206の全て又は当該PAの処理にかかるる情報の複製 216、ポート管理情報 224の全て又は当該PAの処理にかかるるポート管理情報の複製 222、NAS管理情報 223の全て又は当該PAの処理にかかるるNAS管理情報の複製 221、ポート 141または転送制御部 144から受信したデータもしくはコマンドを振り分ける要求振り分け処理プログラム 251、LUパス定義処理プログラム 252、外部デバイス定義処理プログラム 253、NAS／ポート障害処理プログラム 231及びファイルアクセス処理プログラム 229が格納されている。

【0057】

CAの制御メモリ155には、下位論理デバイス管理情報201、キャッシュ管理情報203、物理デバイス管理情報202が格納される。メモリ153には、下位論理デバイス管理情報201の複製211、物理デバイス管理情報202の複製212、外部デバイス管理情報205の全て又は当該CAの処理にかかる情報の複製215、PAまたはAAから受信した入出力コマンドを処理するコマンド処理プログラム254、論理デバイス定義処理プログラム255、ライトアフタ処理プログラム及び外部デバイス接続変更処理プログラム256が格納されている。

【0058】

MAの制御メモリ164には、上位論理デバイス管理情報204、外部デバイス管理情報205、LUパス管理情報206、NAS管理情報223及びポート管理情報224が格納される。メモリ163には、全デバイス管理情報(即ち、上位論理デバイス管理情報204、下位論理デバイス管理情報201、物理デバイス管理情報202及び外部デバイス管理情報205)の複製210、論理デバイス定義処理プログラム255、外部デバイス接続変更処理256、LUパス定義処理252、NAS/ポート障害処理プログラム231、外部デバイス定義処理プログラム253、NAS管理情報変更処理プログラム225及びポート管理情報変更処理プログラム226が格納されている。

【0059】

AAのメモリ149には、NAS管理情報223の全て又は当該AAの処理にかかる情報の複製227、外部デバイス管理情報205の全て又は当該AAの処理にかかる情報の複製215、当該AAのキャッシュ管理情報、NAS/ポート障害処理プログラム231及びファイルアクセス処理プログラム229が格納されている。

【0060】

なお、本実施形態では、全ての管理情報(下位論理デバイス管理情報201、物理デバイス管理情報202、上位論理デバイス管理情報204、外部デバイス管理情報205、LUパス管理情報206、NAS管理情報223、ポート管理情報

224及び各複製)は、ストレージ130の工場出荷時にあらかじめ、初期値が設定されているものとする。その後、適宜、ストレージ管理者または、管理サーバからの制御により後述の方法を用いて値が設定変更され、運用される。

【0061】

図15は、STの制御メモリに格納される、制御情報とストレージ制御処理のためのプログラムの一例を示したソフトウェア構成図である。メモリ193には、全デバイス管理情報(即ち、上位論理デバイス管理情報204、下位論理デバイス管理情報201、物理デバイス管理情報202及び外部デバイス管理情報205)の複製210、NAS管理情報223の複製231、ポート管理情報224の複製232、論理デバイス定義処理プログラム255、LUパス定義処理プログラム252、外部デバイス定義処理プログラム253、NAS管理情報変更処理プログラム225、ポート管理情報変更処理プログラム226及びNAS/ポート障害処理プログラム231が格納される。

【0062】

図3は、上位論理デバイス管理情報203の一構成例を示す図である。上位論理デバイス管理情報は、各上位論理デバイス毎に、上位論理デバイス番号から対応外部デバイス番号までの情報の組を保持するエントリ31～39を有する。なお、切替進捲ポインタを使用しない場合はエントリ39を有しなくても良い。

【0063】

エントリ31には、対応する上位論理デバイスを識別するために上位論理デバイスにMAが割り当てた番号が格納される。エントリ32には、上位論理デバイス番号により特定される上位論理デバイスの記憶容量が格納される。エントリ33には、当該上位論理デバイスに対応付けられている下位論理デバイスの番号と当該下位論理デバイスが属するCAの番号が格納される。上位論理デバイスが未定義の場合、エントリ33には無効値が設定される。尚、この下位論理デバイス番号が、当該下位論理デバイスを管理するCAが保持する下位論理デバイス管理情報201のエントリ番号となる。

【0064】

エントリ34には、対応する上位論理デバイスの状態を示す情報が設定される

。状態としては、「オンライン」、「オフライン」、「未実装」及び「障害オフライン」が存在する。「オンライン」は、当該上位論理デバイスが正常に稼動し、ホスト100またはAAからアクセスできる状態であることを示す。「オフライン」は、当該上位論理デバイスは定義され、正常に稼動しているが、LUパスが未定義であるなどの理由で、ホスト100またはAAからはアクセスできない状態にあることを示す。「未実装」は、当該上位論理デバイスが定義されておらず、ホスト100またはAAからアクセスできない状態にあることを示す。「障害オフライン」は、当該上位論理デバイスに障害が発生してホスト100またはAAからアクセスできないことを示す。

【0065】

なお、本実施形態では簡単のため、ストレージ130の工場出荷時にあらかじめ、ディスク装置157で作成された物理デバイスへ下位論理デバイスが割り当てられ、更に下位論理デバイスに上位論理デバイスが割り当てられているものとする。このため、下位論理デバイスに割当済みの上位論理デバイスについてのエントリ34の初期値は「オフライン」、その他の上位論理デバイスについては上位論理デバイスが定義されていないことになるので、エントリ34の初期値は「未実装」となる。

【0066】

エントリ35には、ポート番号、ターゲットID及びLUNが登録される。ここで登録されるポート番号は、対応する上位論理デバイスが複数のポート141またはAAの転送制御部146のうちどのポートまたはAAの転送制御部146に接続されているかを表す情報、即ち対応する上位論理デバイスにアクセスするために用いられるポート141またはAAの転送制御部146の識別情報である。ここでポート141又はAAの転送制御部146の識別情報は、各ポート141及びAAの転送制御部146に割り振られているストレージ130内で一意な番号であり、エントリ35には、当該上位論理デバイスがLUN定義されているポート141又はAAの転送制御部146の番号が記録される。

【0067】

また、ターゲットIDとLUNは、当該上位論理デバイスを識別するための識

別子である。本実施形態においては、上位論理デバイスを識別するための識別子として、SCSIプロトコルでホスト100またはAAからデバイスをアクセスする場合に用いられるSCSI-ID及び論理ユニット番号（LUN）を用いる。

【0068】

エントリ36には、当該上位論理デバイスへのアクセスが許可されているホスト100またはAAを識別するホスト名が登録される。ホスト名としては、ホスト100のポート107に付与されたWWN(World Wide Name)やiSCSIネームなど、ホスト100又はポート107及びAAを一意に識別可能な値であれば何を用いてもよい。ストレージ130は、このほかに、各ポート141のWWNやMAC(Media Access Control)アドレス、IPアドレスなどの属性に関する管理情報を保持する。

【0069】

エントリ37には、当該上位論理デバイスへの入出力要求の処理形態を示す「CA経由」、「PA直結」又は「PA直結移行中」という値が格納される。

「CA経由」は、当該上位論理デバイスが外部デバイスと対応付けられていない場合や、当該上位論理デバイスが外部デバイスに対応付けられていて、かつ当該上位論理デバイスに対する入出力要求に対してCA150での入出力処理が必要な処理形態を示す。

【0070】

「PA直結」は、当該上位論理デバイスが外部デバイスに対応付けられており、かつ、CAを経由せずに直接PA間で当該上位論理デバイスへの入出力要求を転送できる処理形態を示す。

「PA直結移行中」は、外部デバイス接続形態を「CA経由」から「PA直結」へ切り替える過渡状態、すなわち、後述する切替処理が対応CAにて実行中であることを示す。

【0071】

エントリ38には、当該上位論理デバイスが外部デバイスに対応付けられている場合にはその外部デバイスの識別番号が、そうでない場合は無効値が設定され

る。

【0072】

エントリ39には、切替進捲ポインタが登録される。切替進捲ポインタは、当該上位論理デバイスが「PA直結移行中」である場合に使用され、当該上位論理デバイスが有する記憶領域のうち、外部接続の接続形態変更処理が未完了な領域の先頭アドレスを示す情報である。なお、「PA直結移行中」処理を行わない場合には、エントリ39を省いてもかまわない。

【0073】

図4は、LUパス管理情報206の一構成例を示す図である。LUパス管理情報206は、PAが有する各ポート141又はAAの転送制御部146に対して定義されている有効なLUN毎に、情報を保持する為のエントリ41～43を有する。エントリ41には、ポート141又はAAの転送制御部146に定義された(割り当てられた)LUNのアドレスが格納される。

【0074】

エントリ42には、当該LUNが割り当てられている上位論理デバイスの番号が格納される。エントリ43には、当該LUNに対してアクセスを許可されているホスト100またはAAを示す情報、具体的にはWWNや転送制御部146の番号等が登録される。

【0075】

尚、一つの上位論理デバイスに対して複数のポート141またはAAの転送制御部146が定義され(割り当てられ)ており、複数のポート141またはAAから当該上位論理デバイスにアクセスできる場合がある。この場合、当該複数のポート141またはAAのLUN各々に関するLUパス管理情報206のエントリ43に登録された値の和集合が、当該上位論理デバイスに関する上位論理デバイス管理情報203のエントリ36に保持される。

【0076】

図5は、下位論理デバイス管理情報201の一構成例を示す図である。各CAは、自CAに属する各下位論理デバイス毎に、下位論理デバイス番号から対応上位論理デバイス番号までの情報の組を登録するエントリ51～56を有する。な

お、切替進捲ポインタが使用されない場合は、エントリ 56 は省いてもかまわない。

【0077】

エントリ 51 には、対応する下位論理デバイスを識別するための識別番号が登録される。エントリ 52 には、エントリ 51 により特定される下位論理デバイスの記憶容量が格納される。

【0078】

エントリ 53 には、当該下位論理デバイスに対応付けられる、CA の物理デバイスの識別番号又は外部ストレージ 180 に存在する外部デバイスの識別番号が格納される。当該下位論理デバイスに対して物理デバイス又は外部デバイスが割り当てられていない場合、エントリ 53 には無効値が設定される。なお、エントリ 53 に登録されるデバイス番号は、当該下位論理デバイスを管理している CA が保持する物理デバイス管理情報 202、もしくは MA が保持する外部デバイス管理情報 205 のエントリ番号と一致する。

【0079】

エントリ 54 には、当該下位論理デバイスの状態を示す情報が設定される。設定される情報の内容は、上位論理デバイス管理情報 203 のエントリ 34 と同じであるため、説明は省略する。エントリ 55 には、当該下位論理デバイスに対応付けられている上位論理デバイス番号が設定される。

【0080】

エントリ 56 には、切替進捲ポインタが登録される。このポインタは上位論理デバイス管理情報 204 の切替進捲ポインタと同じで、当該下位論理デバイスに対応する上位論理デバイスが「PA 直結移行中」である場合に使用され、移行中の上位論理デバイスに対応する下位論理デバイスのうち、外部接続の接続形態変更処理が未完了な記憶領域の先頭アドレスを示す情報である。

【0081】

図 6 は、CA のディスク装置 157 から構成される物理デバイスを管理するための物理デバイス管理情報 202 の一構成例を示す図である。各 CA は、自 CA が有する物理デバイス毎に、物理デバイス番号からディスク内サイズの情報の組

を保持する為のエントリ61～69を有する。

【0082】

エントリ61には、物理デバイスを識別するための識別番号が登録される。エントリ62には、エントリ61に登録された値で特定される物理デバイスの記憶容量が格納される。エントリ63には、当該物理デバイスに対応付けられる下位論理デバイスの番号が格納される。当該物理デバイスが下位論理デバイスへ割り当てられていない場合、エントリ63には無効値が設定される。

【0083】

エントリ64には、当該物理デバイスの状態を示す情報が設定される。状態としては、「オンライン」、「オフライン」、「未実装」及び「障害オフライン」が存在する。「オンライン」は、当該物理デバイスが正常に稼動し、下位論理デバイスに割り当てられている状態であることを示す。「オフライン」は、当該物理デバイスは定義され正常に稼動しているが、下位論理デバイスに未割り当てであることを示す。「未実装」は、当該物理デバイスがディスク装置157に対応付けされていない状態にあることを示す。「障害オフライン」は、当該物理デバイスに障害が発生して下位論理デバイスに割り当てられないことを示す。

【0084】

なお、本実施形態では、簡単のため、物理デバイスは製品の工場出荷時にあらかじめディスク装置157に対応付けられて作成され、かつ下位論理デバイスにも割り当てられているものとする。このため、利用可能な物理デバイスについてはエントリ64の初期値は「オンライン」状態、その他は「未実装」状態となる。

【0085】

エントリ65には、当該物理デバイスが割り当てられたディスク装置157のRAIDレベル、データディスクとパリティディスク数などRAID構成に関連する情報が保持される。同じように、エントリ66には、RAIDにおけるデータ分割単位(ストライプ)長が保持される。エントリ67には、当該物理デバイスが割り当てられたRAIDを構成する複数のディスク装置157各々の識別番号が保持される。このディスク装置157の識別番号は、CA内でディスク装置1

57を識別するために付与された一意な値である。

【0086】

エントリ68とエントリ69には、当該物理デバイスが各ディスク装置157内のどの領域に割り当てられているかを示す情報（開始オフセット及びサイズ）が登録される。本実施例では簡単のため、全物理デバイスについて、RAIDを構成する各ディスク装置157内のオフセットとサイズを統一している。

【0087】

図7は、ストレージ130に接続された外部ストレージ180の外部デバイスを管理するための、外部デバイス管理情報205の一構成例を示す図である。ストレージ130のMAは、外部ストレージ180が有する外部デバイス毎に、外部デバイス番号からターゲットポートID／ターゲットID／LUNリストまでの情報の組を保持する為のエントリ71～79を有する。

【0088】

エントリ71には、ストレージ130のMAが当該外部デバイスに対して割り当てた、ストレージ130内で一意な値が格納される。エントリ72には、エントリ71に登録された値より特定される外部デバイスの記憶容量が格納される。

【0089】

エントリ73には、当該外部デバイスに対応付けられているストレージ130内の上位論理デバイスの番号が登録される。エントリ74には、当該外部デバイスに対応付けられ、当該外部デバイスに対するホスト100からのアクセス要求を処理する、ストレージ130のCAの番号と、当該外部デバイスに対応付けられている下位論理デバイスの番号が格納される。また、AAが直接当該外部デバイスを使用する場合は、CAの番号ではなく、AAの番号が格納される。なお、当該外部デバイスに対して上位論理デバイスもしくは下位論理デバイスが割り当てられていない、もしくはAAが直接使用しない場合、エントリ73もしくはエントリ74には無効値が設定される。

【0090】

エントリ75には、当該外部デバイスの状態を示す情報が設定される。各状態の種類は物理デバイス管理情報202のエントリ64と同じである。ストレージ

130は初期状態では外部ストレージ180を接続していないため、エントリ75の初期値は「未実装」となる。エントリ76には、当該外部デバイスを有する外部ストレージ180の識別情報が登録される。外部ストレージ180の識別情報としては、外部ストレージ180のベンダ識別情報と各ベンダが一意に割り振る製造シリアル番号の組み合わせなどが考えられる。

【0091】

エントリ77には、当該外部デバイスに対して当該外部デバイスを有する外部ストレージ180が割り振ったデバイスの識別番号が格納される。なお、外部デバイスは外部ストレージ180の論理デバイスであるから、エントリ77には外部ストレージ180における論理デバイス番号が格納される。

【0092】

エントリ78には、当該外部デバイスへアクセス可能なストレージ130のポート141に割り振られた識別番号と、当該ポート141が属するPAの識別番号が登録される。複数のポート141から当該外部デバイスへアクセスできる場合には、複数のポート識別番号と複数のPA識別番号が登録される。

【0093】

エントリ79には、当該外部デバイスが外部ストレージ180の1つ以上のポート181にLUN定義されている場合、それらのポート181のポートIDおよび当該外部デバイスに割り当てられたターゲットID/LUNが1つ又は複数個保持される。なお、ストレージ130のPAが外部デバイスにアクセスする場合（PAから外部デバイスに対する入出力要求を送信する場合）には、当該外部デバイスが属する外部ストレージ180によって当該外部デバイスに割り当てられたターゲットID及びLUNが、当該外部デバイスを識別するための情報として用いられる。

【0094】

図16は、NAS管理情報223の一構成例を示す図である。NAS構成情報223は、各AAごとに、AA番号～代替AA番号の情報の組を登録するエントリ1601～1606を有する。

エントリ1601には、ストレージ130内でAAを識別するための識別情報

が登録される。

【0095】

エントリ1602には、ホスト100がファイルアクセス用の通信コマンドを送信する際に送信先を指定するためのIPアドレス、当該AAが管理するファイルが存在するディレクトリ名、当該ディレクトリ名で指定されるディレクトリの情報とディレクトリ下に存在するファイルを格納するために割り当てられている上位論理デバイスの番号もしくは外部デバイス番号、及び当該上位論理デバイスをAAに対して提供しているCAの番号の組からなる情報のリストが登録される。なお、IPアドレスのみを登録し、登録されたIPアドレス先に来た全要求を当該AAが処理するようにしても良い。

【0096】

エントリ1602に複数のIPアドレスが登録されている場合、ホスト100からは、複数のNASがシステム内にあるかのように見える。

また、1つのIPアドレスに複数のAAを割り当てて、例えばディレクトリ毎に異なるAAが処理するようにすることもエントリ1602の設定で可能である。また、AAが複数の上位論理デバイスを例えばディレクトリ毎に使い分ける場合もエントリ1602の設定で可能である。

【0097】

エントリ1603には、AAの処理状態を示す情報が登録される。登録される情報には、「オフライン」、「障害」及び「オン中」の3つが有る。

「オフライン」とは、AAが初期設定処理中だったり、他AAからの情報引継ぎ処理中であってホスト100からの要求を受け付けない状態を示す。

「障害」とは、AAに障害が発生した状態を示す。

「オン中」とは、AAがホスト100からの要求を受け付け可能な状態を示す。

【0098】

エントリ1604には、AAのアクセスモードを示す情報が登録される。アクセスモードを示す情報とはAAが他の部位（CAやPA）と連携して動作を行うかどうかを示す情報であり、「連携」と「非連携」の二種類がある。

「非連携」が設定される時には、AAの処理にかかるコマンドおよびデータは、全てAAを経由する。具体的には、ホスト100からPAが受信したコマンドはAAに送信され、AAは、必要に応じ入出力要求をCAもしくはPA経由で外部ストレージ180に送信する。又、その応答コマンドおよびデータはCAもしくはPAからAAに送信され、処理終了後、AAは応答コマンドをホスト100からのコマンドを受信したPAに送信する。

【0099】

「連携」が設定される時には、AAの入出力要求に基づく、CA又はPA経由で接続された外部ストレージ180から送信されるデータはAAを介さず直接ホスト100からコマンドを受信したPAへ送信され、CA又はPA経由で接続された外部ストレージ180へ送信するデータは、ホスト100からコマンドを受信したPAから直接CA又は外部ストレージ180と接続されているPAに対し送信される。これにより、大量のデータがAAを経由しないため、AAの処理負荷と相互接続網170の負荷が軽減される。なお、本実施形態では、簡単のため、ストレージ130の工場出荷時にあらかじめ、初期値として「非連携」がエントリ1604に設定されており、その後、適宜、ストレージ管理者または、管理サーバからの制御により後述の手段を用いて値が設定変更されるものとする。「連携」は、例えばサイズの大きいファイルに対するアクセスが多いと推測される環境において、処理負荷の軽減のためにストレージ管理者により採用される。なお、エントリ1604を省略し、AAとして「連携」もしくは「非連携」のいずれか一方を提供することにしてもよい。

【0100】

エントリ1606には、当該AAに障害が発生した場合に、当該AAの管理情報を引き継ぎ処理を行う代替AAの番号が設定される。代替AAが無い場合には、無効値が設定される。代替AAを設けない場合には、エントリ1606は省略されてもよい。

【0101】

図17は、ポート管理情報224の一構成例を示す図である。ポート管理情報224は、ポート141がIPネットワークと接続される場合にその接続状態を管

理する際にMA等の各部位で用いられる。尚、ポート141がFCネットワークに接続される場合には、そのポート141は、ポート管理情報224により管理されない。ポート管理情報224は、IPネットワークと接続されるポート141ごとに、そのポートに関する情報を登録するエントリ1701～1708を有する。

【0102】

エントリ1701には、ポート141をストレージ130内で一意に識別するために割り付けられた番号が設定される。エントリ1702には、当該ポート141のMACアドレスが登録される。エントリ1703には、当該ポート141に割り当てられたIPアドレスのリストが登録される。エントリ1704には、当該ポート141が処理できるプロトコル（例えば、iSCSI、NFS等）のリストが登録される。ポート141は、自ポートのエントリに登録されたMACアドレス1702、IPアドレス1703、プロトコル1704に従い通信処理を行う。なお、処理できないプロトコルのデータをポート141が受信した場合には、当該データは、ポート141にて廃棄処理される。

【0103】

エントリ1705には、当該ポート141の状態が登録される。状態には「通信可能」、「ネットワーク非接続」、「通信不可能」及び「障害」の四種類がある。

「通信可能」は、ポート141に障害が無くかつネットワークに接続されており、ホスト100との通信が可能な状態にあることを示す。「ネットワーク非接続」は、ポート141にケーブル等が接続されておらずネットワークが接続されていない状態を示す。「通信不可能」は、ポート141がネットワークに接続されているが、ネットワークの問題でホスト100と通信ができない状態にあることを示す。「障害」は、ポート141に障害ある状態を示す。

【0104】

エントリ1707には、ポート141が外部ストレージ180に接続する場合に、当該ポート141に接続された外部ストレージ180の外部デバイス番号が登録される。ポート141が外部ストレージ180に接続されていない場合には

、無効値が設定される。

【0105】

エントリ 1708 には、当該ポート 141 に障害が発生した場合に、当該ポート 141 に割り当てられたIPアドレスのリスト 1703 を引継ぎ当該IPアドレスに対するアクセスを続行するポート 141（以下「代替ポート」）のポート番号が登録される。なお、代替ポートが通信に使用されていなかった場合に引継ぎ処理を行う場合、IPアドレスのリストだけでなく、MACアドレスも代替ポートへ引き継いでもかまわない。

【0106】

なお、ポート管理情報 224 でFCプロトコルに対応するするポートも管理する場合、前述の情報に加え、エントリ 1702 にはWWNが、エントリ 1703 にはポートIDが、エントリ 1704 にはFCを示す情報が登録される。

【0107】

以下、ホスト 100 がストレージ 130 に対してファイルアクセスを行った場合のストレージ 130 内部での処理手順について説明する。ファイルアクセスの処理は、具体的には、要求振り分け処理、ファイルアクセス処理、連携モード処理、コマンド処理及びライトアフタ処理の各処理から構成される。以下、順に説明する。

【0108】

図 11 は、要求振り分け処理 251 の処理フローの一例を示す図である。要求振り分け処理 251 は、PA のポート 141 においてホスト 100 または外部ストレージ 180 から受け取った入出力要求やデータ、および転送制御部 144 において CA、AA 又は別の PA から受け取った入出力要求やデータやコマンドを、CA、AA、他の PA、ホスト 100 又は外部ストレージ 180 等、適切な対象部位へ PA が振り分ける処理である。

【0109】

ポート 141 又は転送制御部 144 からデータやコマンドなどを含んだフレームを受信した PA は（ステップ 1101）、フレーム内のプロトコルの種別および当該プロトコルのコマンドの種別によって処理を切り分ける。なお、プロトコ

ルの種別は、当該ポート141がサポートするネットワーク（IPもしくはFC）と受信したフレーム内のデータから判別することができる（ステップ1102）。

【0110】

当該フレームがFCPコマンドフレームの場合もしくはiSCSIコマンドであった場合、PAは、当該フレームの送信元によって処理を切り分ける。なお、本実施形態では、コマンドフレームは、ポート141を介してホスト100から受信する場合と、転送制御部144を介して外部ストレージ180をアクセスするCA、AA又は他PAから受信する場合の2つがある（ステップ1103）。

【0111】

まず、送信元がホスト100の場合、PAは、受信したコマンドのヘッダ情報に含まれるアクセス対象の上位論理デバイスを表すLUNに基づいて、LUバス管理情報の複製216及び上位論理デバイス管理情報の複製214を参照し、当該上位論理デバイスが外部デバイスに対応付けられており、かつデバイスアクセスモードが「CA経由」でないかを判断する。当該上位論理デバイスが外部デバイスに対応付けられておりかつデバイスアクセスモードが「PA直結」である場合、PAは更に、当該コマンドの要求種別がCAでの処理不要であるかを判断する。これらの条件を全て満たす場合、PAは、当該要求のアクセス形態はPA直結であると判断し、いずれかの条件を満たさない場合は当該要求のアクセス形態はCA経由であると判断する（ステップ1104～1105）。

【0112】

ステップ1105でアクセスモードが「CA経由」でなく、コマンド要求種別がCA処理不要の場合、PAはさらにアクセスモードが「PA直結移行中」でかつ当該コマンド内LBAとブロック数で示されるアクセス領域が当該上位論理デバイスの切替進捲ポインタよりも後にあるかを判断する（ステップ1119）。アクセスモードが「PA直結移行中」で無い又はPAが受信したコマンドが進捲ポインタよりも前、すなわちPA直結へ切替済みの領域を対象としている場合には、PAは、コマンドのアクセス先を外部デバイスを有する外部ストレージ180が接続されたPAとする（ステップ1106）。

【0113】

ステップ1119でアクセスモードが「PA直結移行中」でかつアクセス先が未切替な部分であると判断された場合及びステップ1105でCA経由アクセスであると判断した場合、PAは上位論理デバイス管理情報の複製214を参照して、受信したコマンドのアクセス対象となっている上位論理デバイスが対応するCA番号などを算出し、これを入出力処理の要求先、当該フレームの送信元を入出力処理の要求元として、当該コマンドのルーティング制御情報として記憶する（ステップ1107）。

【0114】

なお、PA直結移行中処理を行わない場合、ステップ1119は省略してもかまわない。この場合、図3のエントリ39は省略される。

【0115】

一方、ステップ1103の切り分け処理で、当該フレームの送信元がホスト100以外、すなわちCA、AA又は他PA140の場合、PAは当該フレームのヘッダ情報で指定されている外部ストレージ180の外部デバイスを転送先に決定し、これを当該入出力処理の要求先、当該フレームの送信元を入出力処理の要求元として当該コマンドのルーティング制御情報を登録する（ステップ1108、1109）。

【0116】

一方、ステップ1102の切り分け処理で、PA140で受信したフレームの種別がデータフレーム、受信準備完了通知フレーム又は完了報告フレームの場合、PAは、当該フレームに対応する入出力処理のルーティング制御情報を参照し、当該フレームの送信元部位の交信相手先を当該フレームの転送先として決定する。すなわち、PAが当該フレームを外部ストレージ180から受信した場合は、当該入出力処理の要求元であるホスト100が接続されたPA、CA又はAAがフレームの転送先となる。又、PAが当該フレームをホスト100から受信した場合は、当該入出力処理の要求先であるCA又は外部ストレージ180が接続されたPAが当該フレームの転送先として決定される。さらに、PAが当該フレームをCA又は他PAから受信した場合は、当該入出力処理の交信相手先（即ち当該フレームの送信元が当該入出力処理の要求元ならば要求先）であるホスト1

00又は外部ストレージ180が、それぞれ当該フレームの転送先として決定される（ステップ1110）。

【0117】

ステップ1106、1107、1109及び1110で受信フレームの転送先が決まつたら、PAは転送先部位によって処理を切り分ける（ステップ1111）。

転送先が他のPAの場合、PAは、当該入出力要求のルーティング制御情報に登録されている転送先のアドレスを当該フレームの送信先アドレスへ設定する。すなわち、送信先が外部デバイスの場合、PAは、受信したフレームのデスティネーションIDをルーティング制御情報に当該入出力処理の要求先として登録されている外部デバイスのターゲットポートID（即ちポート181のポートID）へ、また、同フレーム内のLUNなどのアドレス情報を当該外部デバイスのアドレス（即ち外部デバイス管理情報を参照して取得したLUN等）に書き換える。送信先がホスト100の場合、PAは、受信フレームのデスティネーションIDをルーティング制御情報に当該入出力処理の要求元として登録されているホスト100のポート107のポートIDへ書き換える（ステップ1112）。

【0118】

また、転送先が当該PA内のポート141に接続されたホスト100又は外部ストレージ180の場合は、PAは、当該ポート141のアドレスを当該フレームの送信元アドレスへ設定する（ステップ1113）。

そして、コマンド、データ、受信準備完了報告、完了報告フレームを受信したPAは、決定された転送先であるホスト100、又は外部ストレージ180、CA、AA又は他のPAへ当該フレームを転送する（ステップ1114）。さらに、転送を完了した当該フレームが完了報告フレームだった場合は、PAは、当該入出力処理のルーティング制御情報の登録を解除する（ステップ1115、1116）。

【0119】

また、ステップ1102で、PAが受信したフレームがNASのコマンド、即ちファイルアクセスのコマンドであった場合、PAは、フレームヘッダにセットさ

れているIPアドレスおよびコマンドにセットされているファイル名から得られるディレクトリ情報を元に、NAS管理情報の複製221を検索し、コマンドを処理するべきAAを決定し、当該AAにコマンドを転送する。つぎに、AAはファイルアクセス処理229（ステップ1118）を行う。尚、ファイルアクセス処理229の詳細については、後述する。

【0120】

また、ステップ1102で、PAが受信したフレームがFCPもしくはiSCSI関連のフレームでなく、かつ、NASのコマンドでもなかった場合は、フレームがFCのフレームであればファイバチャネルのノードポートとして従来公知な処理を当該フレーム受信PAの制御プロセッサ142にて実行し、フレームがIPのフレームであればPAはフレームの廃棄処理を行う（ステップ1117）。

【0121】

なお、要求振り分け処理が実行された結果、コマンドの転送先がCA又はAAであると決定された場合、PAは、フレーム以外にアクセス対象であるCA番号またはAA番号、下位論理デバイス番号または上位論理デバイス番号または外部デバイス番号、LBA、サイズなどの情報を付加して転送先のCAもしくはAAへ送信する。また、コマンドに続くデータなどをCAもしくはAAに転送する場合、PA、CA又はAAのそれぞれの転送制御部において、既に確立したコマンドシーケンスを意識したデータ転送を行う。

【0122】

図21は、ファイルアクセス処理229の処理フローの一例を示す図である。要求振り分け処理251の結果、AAはPAからコマンドを受信する（ステップ2101）。AAは、コマンドで要求された処理対象のファイル関連の情報が自身のメモリ149にキャッシュされているかをキャッシュ管理情報230でチェックする（ステップ2102）。処理対象のファイル関連の情報（ディレクトリ情報等）が無かった場合、AAはNAS管理情報の複製227を参照して関連情報が記憶されているデバイス（CAが提供する論理デバイス又は外部デバイス）からディレクトリ情報を読み出しメモリ149へ格納する（ステップ2103）。

【0123】

情報があった場合、または、情報を読み出しキャッシュに格納した後、AAは、要求がリードもしくはライトであるか否かをチェックする（ステップ2104）。リード又はライトでなかった場合、AAは既存技術を使い当該コマンドの処理を行う（ステップ2118）。

【0124】

要求がリード又はライトであった場合、AAはデータが自己のメモリ149のキャッシュ領域にあるか否かチェックする（ステップ2105）。要求がリードで、かつ既にデータがメモリ149のキャッシュ領域にあった場合、AAはそのデータを元にリード処理を行う。

要求がリードで、かつデータがメモリ149のキャッシュ領域に無かった場合、AAはメモリ149にキャッシュ領域を確保する（ステップ2106）。キャッシュ領域を確保後、もしくは、要求がライトであった場合、AA自身のアクセスモードが「連携」か否かをチェックする（ステップ2107）。

【0125】

「連携」でなかった場合、AAはコマンドの内容とディレクトリ情報に従い、当該データが記録されているCA又は外部ストレージ180が接続されているポート141があるPAへ入出力要求を送信する（ステップ2108）。送信したコマンドがリード要求であった場合（ステップ2109）、AAは、コマンド処理254を実施したCAもしくは要求振り分け処理251を実施したPAからデータを受信し、受信したデータをキャッシュ領域に格納し、その後完了報告を受信（ステップ2110、2113）する。

【0126】

ステップ2109で送信したコマンドがライト要求であった場合、AAは、コマンド処理254を実施したCAもしくは要求振り分け処理251を実施したPAから準備完了報告を受信し（ステップ2111）、ついでデータを送信し（ステップ2112）、完了報告を受信する（ステップ2113）。

その後、AAは受信したデータに基づいて、コマンドで要求された操作ができる分のデータのリードを完了したか、またはデータのライトを完了したかをチェックし、完了していなかった場合、AAはステップ2101～2114の処理を

データのリードまたはライトが完了するまで繰り返す（ステップ2114）。

ステップ2107で連携モードであった場合、AA、PA、SAは、連携モード処理を行う（ステップ2117）。

【0127】

ステップ2105で処理を完了した場合、ステップ2114、ステップ2117及びステップ2118の処理の終了後、AAは、その結果をコマンド応答にしてコマンド送信元のPAへ送信する（ステップ2115）。コマンド応答を受信したPAは、コマンド送信元のホスト100へ当該コマンド応答を送信する（ステップ2116）。

【0128】

なお、本実施形態では、PAでは、TCP/IPまでの処理を行い、NFS等NASのアクセスに使用されるプロトコルにしたがったコマンドがPAとAA間で送受される。

また、本実施形態では、ライト要求に伴うデータを直接CAに書き込む形をとったが、一旦AAのメモリ149のキャッシュ領域に書き込んでから、適当な時にCAに書き込む方式でも良い。また、本実施形態では、AAが、1回以上のリード／ライト要求をCA又は外部ストレージ180に接続されるPAに対して行う方式をとったが、CA、PA及びAAとの間で指示方法を予め決めておき、一回のリード／ライト要求で処理が終了するような方式としても良い。

【0129】

図22は、ファイルアクセス処理229の処理内で実施される連携モード処理2117の処理フローの一例を示す図である。連携処理の場合、AAは、コマンド送信元のPAへ連携モードで動作することを通知する（ステップ2201）。次にAAは、リードまたはライトを行うデバイスの処理を行うCAまたは外部ストレージ180と接続されるポート141があるPAへ連携モードでのリード／ライト要求を送信する（ステップ2202）。

【0130】

送信したのがリード要求であった場合（ステップ2203）、コマンド送信元のPAは、コマンド処理254を実施したCA又は要求振り分け処理251を実

施したPAからデータを受信（ステップ2204）し、AAは完了報告をコマンド処理254を実施したCA又は要求振り分け処理251を実施したPAから受信（ステップ2205）する。ステップ2203でライト要求であった場合、コマンド送信元のPAは、コマンド処理254を実施したCA又は要求振り分け処理251を実施したPAから準備完了を受信し（ステップ2207）、ついでデータを送信する（ステップ2208）。その後、AAは、その完了報告をコマンド処理254を実施したCA又は要求振り分け処理251を実施したPAから受信する（ステップ2205）。

【0131】

AAは受信したデータでコマンドで要求された操作ができる分のデータのリードを完了したか、またはデータのライトを完了したかをチェックし（ステップ2206）、完了していなかった場合は、ステップ2202～2206の処理をデータのリードまたはライトが完了するまで繰り返す。

【0132】

データリードまたはライトが完了した後、AAは図21の処理フローに戻り、結果をコマンド応答にしてコマンド送信元のPAへ送信し（ステップ2115）、PAは、先にAAから受信した連携モード通知、PAまたはCAから受信したデータ、AAから受信したコマンド応答からコマンド応答を作成し、コマンド送信元のホスト100へ当該コマンド応答を送信する（ステップ2116）。

【0133】

図12は、コマンド処理254の処理フローの一例を示す図である。コマンド処理は、リードやライトの要求を受信したCAにおいて下位論理デバイスに対する入出力要求を処理するものである。

【0134】

要求振り分け処理251の結果又はファイルアクセス処理229の結果、PA又はAAから受信したコマンドについて、CAは、コマンド種別をチェックする（ステップ1202）。当該コマンドの要求がリードの場合、CAは、ディスクキャッシュ154上にリード対象のデータがヒットしているか（キャッシュに格納されているか）をキャッシュ管理情報203を参照して判定する（ステップ1

203)。

【0135】

キャッシングの場合、CAはキャッシング管理情報203を更新してディスクキャッシング154に領域を確保し（ステップ1204）、当該データが格納される物理デバイス又は外部デバイスからキャッシングへデータを読み出す。リードデータが格納されているデバイスが物理デバイスであるか外部デバイスであるかは、CAが下位論理デバイス管理情報210を参照することによって判断することができる（ステップ1205）。

【0136】

リードデータが格納されているデバイスが物理デバイスであれば、CAは物理デバイス管理情報202より特定したディスク装置157に対してリード要求を発行してデータを読み出し（ステップ1206）、読み出したデータをディスクキャッシング154の確保領域に格納する（ステップ1207）。

【0137】

リードデータが格納されているデバイスが外部デバイスであれば、CAは外部デバイス管理情報205を参照してエントリ78からアクセス対象PAを特定し、外部デバイスに対するリード要求を当該PAへ送信する（ステップ1210）。

リード要求を受信したPAでは要求振り分け処理251によって当該リード要求を外部ストレージ180に対して送信し、応答として外部ストレージ180からリードデータを受信し、外部ストレージ180から受信したリードデータをCAへ返送する（ステップ1211）。CAは、PAから受信したリードデータをディスクキャッシング154の確保領域に格納する（ステップ1207）。

【0138】

ディスクキャッシングにデータが存在していた又はディスクキャッシングにリードデータが格納されたら、CAは、ディスクキャッシング154に格納されたデータをコマンド転送元のPA又はAAへ送信し（ステップ1208）、PAは要求振り分け処理251を実行して当該データをホスト110へ転送し、AAはファイルアクセス処理の続きをを行う。

【0139】

CAがPA又はAAから受信したコマンドがライトの場合、CAはライトデータに対応する旧データがディスクキャッシュ154に格納されているか否かを判断し（1212）、キャッシュミスの場合にはディスクキャッシュ154領域を割り当てる（1213）。

【0140】

ステップ1212又は1213の終了後、CAは、コマンド転送元のPA又はAAに対して、転送準備完了を送信する（ステップ1214）。PAへ転送準備完了を送信した場合、転送準備完了を受信したPAは要求振り分け処理251を実行して転送準備完了をホスト100へ転送し、その結果ホスト100からライトデータが送られてくると、PAは転送準備完了を送信したCAへデータを送信する。一方、AAへ転送準備完了を送信した場合、転送準備完了を受信したAAは、CAへデータを送信する。CAではPA又はAAから受信したライトデータを、確保したディスクキャッシュ154上の領域へ格納する（ステップ1215、1216）。

【0141】

また、CAで受信したコマンドがリードでもライトでもない場合、例えばモードセンスなどのセンス系コマンドや診断系のコマンドの場合は、CAは、一般的なSCSIストレージの仕様に基づき適切な処理を実行する（ステップ1217）。

【0142】

ステップ1208、1217又は1216の処理を終了したCAは、コマンド処理を完了したことをコマンド転送元のホスト100に接続されているPA又はAAへ報告する。完了報告を受領したPAは、要求振り分け処理251を実行して、ホスト100へ完了報告フレームを送信する。完了報告を受領したAAは、ファイルアクセス処理の続きをを行う（ステップ1209）。

【0143】

なお、ステップ1201でCAがAAからコマンドを受信した場合でAAが「連携」モードで動作している場合、CAは、AAからはコマンドの他に連携モー

ドの通知と連携するべき PA を指定する情報等を受信する。この場合、CA は、ステップ 1208 およびステップ 1214、1215 における準備完了とデータの送受信を、連携を指示された PA に対して行う。「連携」モードで動作している PA は、AA からの指示と CA からの準備完了およびデータの送受信に従い、連携モード時のファイルアクセス処理を完了する。

【0144】

図 13 は、ライトアフタ処理 257 の処理フローの一例を示す図である。ライトアフタ処理とは、CA におけるコマンド処理 254 の結果、ディスクキャッシュ 154 に格納されたライトデータを CA がディスク装置 157 へ書き出す処理である。ディスクキャッシュ 154 上に保持されたライトデータはキャッシュ管理情報 203 で管理される。通常、ライトデータやディスクからリードされたりードデータは、なるべく古いものから順にディスクキャッシュ 154 より追い出されるように、キューなどで管理されている。

【0145】

CA は、キャッシュ管理情報 203 で管理されたデータの中から、ディスク装置 157 へ実際に書き込むデータを選択し（ステップ 1301）、当該データが格納されるディスク装置 157 が外部デバイスに対応づけられているか物理デバイスに対応付けられているかを下位論理デバイス管理情報 201 に基づいて判断する（ステップ 1302）。

【0146】

ライトデータを物理デバイスに書き込む場合、CA は物理デバイス管理情報 202 を参照してデータを書き込むディスク装置 157 を特定し、ライトデータを特定されたディスク装置 157 へ転送する（ステップ 1303）。

【0147】

一方ライトデータを外部デバイスに書き込む場合には、CA は外部デバイス管理情報 205 を参照してライトデータが書き込まれる外部デバイスを有する外部ストレージ 180 と接続されている PA を特定し、当該 PA へライト要求を送信する（ステップ 1304）。ライト要求を受信した PA は、要求振り分け処理 251 を実行して当該ライト要求を外部ストレージ 180 に転送し、ライト要求に

に対する応答として当該外部ストレージ180から転送準備完了通知を受信する。すると当該PAは転送準備完了通知をライト要求送信元のCAへ送信する。完了通知を受信したCAはライトデータを当該PAへ送信する（ステップ1305、1306）。

【0148】

ライトデータを受信したPAは、要求振り分け処理251を実行してライトデータを上述の外部ストレージ180に送信し、外部ストレージ180によってライトデータが外部デバイスに書き込まれる。

【0149】

ライトデータを物理デバイスもしくは外部デバイスへライトした後、CAは、ライトデータが格納されていたディスクキャッシュ154上の領域を解放する（ステップ1307）。

【0150】

次に、AA又はPAのポート141に障害が発生した場合のストレージ130の動作について説明する。AA又はPAのポート141に障害が発生した場合、ストレージ130は、代替のAA又はポート141を用いて処理を継続する。

【0151】

図20は、NAS／ポート障害処理231の処理フローの一例を示す図である。ポート141もしくはAAの障害を検出または他の部位から通知されたMAは、検出もしくは通知された障害がポート障害かAA障害かを判断する（ステップ2001、2002）。検出等された障害がポート障害であった場合、MAは、障害が発生したポート141のポート管理情報224を参照し、当該ポート141の管理情報のエントリ1705を「障害」に設定する。

【0152】

次に、MAは、エントリ1708に登録されている代替ポートのポート管理情報に障害の発生したポート141の情報を追加コピーする。代替ポートのエントリ1705が「通信」状態の場合は、障害の発生したポート141のエントリ1704に登録されていたIPアドレスリストを代替ポートのIPアドレスリストに追加する。「通信」状態でなかった場合、MAは、障害の発生したポート141の

MACアドレス、IPアドレスリスト、アクセスモード等を代替ポートに対応するポート管理情報の各エントリにコピーする。そして、MAはPAにポート管理情報の更新を指示する（ステップ2003）。

【0153】

AA障害であった場合、MAは障害が発生したAAに対応するNAS管理情報223を参照し、当該管理情報223のエントリ1603を「障害」に設定する。次に、MAは障害が発生したAAに関するNAS管理情報223のエントリ1606に登録されている代替AAのNAS管理情報223に、障害の発生したAAのIPアドレス、ディレクトリ名、対応CA番号、上位論理デバイス／外部デバイスのリストを追加する。その後、MAは、AAとPAにNAS管理情報223の更新を指示する（2004）。

【0154】

MA、PA及びAAでポート管理情報とNAS管理情報の更新が終了した後、MAは、STにポート管理情報とNAS管理情報の更新を通知する（ステップ2005）。ポート管理情報とNAS管理情報の更新を通知されたSTは、更新のあった情報を取り込んだ後、要求元であるストレージ管理者もしくは管理サーバへ報告すべく完了報告を出力する（ステップ2006）。

【0155】

図18は、NAS管理情報変更処理225の処理フローの一例を示す図である。なお、新規にNAS管理情報を登録する場合、および、あるAAにかかるNAS管理情報の一部もしくは全てを別AAにコピーして処理を引き継がせる場合も本処理が実行される。更に本処理は、AAおよびPAが各自の処理負荷を測定する機構を持ち、その処理負荷の情報をMAもしくはST経由で管理サーバが収集し、その情報を元に、ストレージ管理者または、管理サーバ上の記載しない処理負荷管理処理部が処理負荷の高いAAもしくはAAの処理にかかるPAから、処理負荷の低いAAもしくはPAへ負荷を分散させるように制御する時に動作する。

【0156】

また、一部のAAの修理や、AAを高性能なものに変更する等のために新しい物と交換するような場合、ストレージ管理者または、管理サーバからの制御で動

作させる。

【0157】

ストレージ管理者もしくは管理サーバ110からNAS管理情報変更指示を受け付けたS Tは、MAに対して同指示を転送する（ステップ1801）。同指示には、情報を変更するAAにかかるNAS管理情報エントリが付加される。指示を受信したMAは、NAS管理情報を変更すると同時に、変更するNAS管理情報のエントリにかかるAAに対し、NAS管理情報を変更するように指示を送信する（ステップ1802）。指示を受信したAAは、NAS管理情報の複製227の変更を行う（ステップ1803）。変更指示が新たなエントリの追加であった場合（ステップ1804）、MAに登録の完了を通知する（ステップ1807）。

【0158】

変更指示が既存エントリの変更であった場合、AAは、当該変更にかかる処理中のコマンドを廃棄し（ステップ1805）さらに、メモリ149にキャッシュされている当該変更にかかるディレクトリ情報をクリアして（ステップ1806）、MAに登録の完了を通知する（ステップ1807）。AAからの登録完了通知を受信したMAは、NAS管理情報の更新があったことを各PA及びS Tに通知する（ステップ1808）。

【0159】

NAS管理情報更新を通知されたPAでは、更新のあったNAS管理情報の複製221をメモリ143へ取り込む。又、NAS管理情報更新を通知されたS Tでは、更新のあった情報を取り込んだ後、要求元であるストレージ管理者もしくは管理サーバ110へ報告すべく完了報告を出力する（ステップ1809）。

【0160】

尚、NAS／ポート障害処理にてNAS管理情報が変更される際には、上述した処理手順のうち、ステップ1802から処理が開始される。

【0161】

図19は、ポート管理情報変更処理226の処理フローの一例を示す図である。なお、新規にポート管理情報を登録する場合及びあるポート141にかかるポート管理情報の一部もしくは全てを別ポート141にコピーして処理を引き継

がせる場合も、本処理が実行される。更に本処理は、PAが各自の処理負荷を測定する機構を持ち、その処理負荷の情報をMAもしくはST経由で管理サーバが収集し、その情報を元に、ストレージ管理者または、管理サーバ上の記載しない処理負荷管理処理部が処理負荷の高いポートから、処理負荷の低いポートへ負荷を分散さるように制御する時に動作する。また、一部のPAの修理や、PAを高性能なものに変更する等のために新しい物と交換するような場合、ストレージ管理者または、管理サーバからの制御で動作させる。

【0162】

ストレージ管理者もしくは管理サーバ110からNAS管理情報変更指示を受け付けたSTがMAに対して同指示を転送する。同指示には、情報を変更するポートにかかるポート管理情報のエントリが付加される（ステップ1901）。指示を受信したMAは、ポート管理情報を変更すると同時に、変更するポート管理情報のエントリにかかるポート141が存在するPAに対し、ポート管理情報の複製222を変更するように指示を送信する（ステップ1902）。

【0163】

指示を受信したPAは、ポート管理情報の複製222の変更を行う（ステップ1903）。変更指示が新たなエントリの追加または、既存エントリへの情報の付加であった場合（ステップ1904）、PAはMAに登録の完了を通知し、更新後のポート管理情報の複製222の内容に従い自PA内のポート141で通信処理を開始する（ステップ1907）。変更指示が既存エントリの内容の変更であった場合、PAは当該変更にかかる処理をクリアし、更新後のポート管理情報の複製222の内容に従い自PA内のポート141で通信処理を開始した（ステップ1905）後、MAに登録の完了を通知する（ステップ1907）。PAからの登録完了通知を受信したMAは、ポート管理情報の更新があったことをSTに通知する（ステップ1908）。

【0164】

ポート管理情報更新を通知されたSTは、更新のあった情報を取り込んだ後、要求元であるストレージ管理者もしくは管理サーバ110へ報告すべく完了報告を出力する（ステップ1909）。

【0165】

尚、NAS／ポート障害処理にてポート管理情報が変更される際には、上述した処理手順のうち、ステップ1902から処理が開始される。

【0166】

次に、本システムの各部位で動作するストレージ制御処理について説明する。ストレージ130内に存在する物理デバイス、及び外部ストレージ180に存在する外部デバイスを含むデバイスを、特定のホスト100またはAAに割り当てて当該ホスト100またはAAから使用できるようにする処理は、外部デバイス定義処理253、論理デバイス定義処理255、LUパス定義処理252の大きく3つに分けることができる。システム管理者は、上述の3つの処理を順に行うことで、上位論理デバイスをホスト100又はAAに割り当てるようストレージ130を設定することができる。以下、この3つの処理について説明する。

【0167】

図8は、外部デバイス定義処理253の処理フローの一例を示す図である。外部デバイス定義処理253は、ストレージ130の管理下に外部ストレージ180内のデバイスを外部デバイスとして導入するための処理である。

【0168】

まず、ストレージ管理者もしくは管理サーバ110からの外部ストレージ180接続指示を受け付けたSTが、MAに接続指示を送信する。接続指示には、接続対象となる外部ストレージ180を特定する情報、例えば、外部ストレージ180のポート181のWWN、IPアドレス、又は外部ストレージデバイスへInquiryコマンドを送信した際にその応答により得られる装置識別情報もしくはその両方の情報及び外部ストレージ180に接続されるストレージ130のポート141の番号とが付加される（ステップ801）。

【0169】

外部ストレージの接続指示を受信したMAは、接続指示に付加された全てのポート141の番号に対応する全PAへ、外部ストレージ180を識別するための情報と外部ストレージ180と接続されるポート141の番号とを有する外部ストレージ接続指示を送信する（ステップ802）。

【0170】

MAから外部ストレージ接続指示を受信したPAは、接続指示に付加された外部ストレージ180の識別情報（以下「外部ストレージ識別情報」）を用いて、接続対象となる外部デバイスを探索する。具体的には、外部ストレージ識別情報としてポート181のWWNまたはIPアドレスを受信した場合、PAは接続指示に付加されたポート番号により指定されたポート141から外部ストレージ180のポート181に対応する全LUNに対してInquiryコマンドを送信し、Inquiryコマンドに対する正常な応答を返したLUNを外部デバイス登録候補とする。

【0171】

外部ストレージ識別情報として装置識別情報しか受信しない場合、PAは、当該PAの全ポート141各々によって検出された外部ストレージ180のノードポート（ノードポートログイン時に検出済み）のそれぞれに対応する全LUNにInquiryコマンドを送信する。その後、Inquiryコマンドに対する正常な応答を返したデバイス（LUN）について、PAは応答内に含まれる装置識別情報を外部ストレージ識別情報と比較し、一致するデバイスを外部デバイス登録候補とする（ステップ803）。

【0172】

PAは、検出された外部デバイス登録候補についての情報リストをMAに返却する。この情報リストには、各外部デバイス登録候補について外部デバイス管理情報205を設定するのに必要な情報が含まれる（ステップ804）。

【0173】

情報リストを受取ったMAは、受け取った外部デバイスリストに含まれる外部デバイスに採番し、外部デバイス管理情報205へデバイス情報を登録し、同情報に更新のあった旨を他の各部位へ通知する。なお、本処理ステップにおいてMAが外部デバイス管理情報205へ登録する情報には、具体的には、MAが外部デバイスに対して割り当てた外部デバイス番号、Inquiryコマンドに対する応答から取得されるサイズ、ストレージ識別情報及び外部ストレージ内デバイス番号、PAからMAに通知されるPA/イニシエータポート番号リスト及びタ

ーゲットポートID/ターゲットID/LUNリストが含まれる。

【0174】

また、対応上位論理デバイス番号及び対応CA番号/下位論理デバイス番号/A番号は未割当のため、MAは初期値である無効値を対応するエントリに設定する。更にエントリ75には「オフライン」状態が設定される（ステップ805）。

【0175】

外部デバイス管理情報205の更新通知を受けた各CA、PA及びAAは、MAの制御メモリ164に保存されている外部デバイス管理情報205を各自のメモリに読み込む。また、STは外部デバイス管理情報205をメモリに取り込むと共に、外部デバイス定義処理の完了を、当該処理の要求元であるストレージ管理者もしくは管理サーバ110に通知すべく完了通知を出力する（ステップ806）。

【0176】

なお、ストレージ130に対してストレージ管理者もしくは管理サーバ110が接続指示と共に、導入対象となる外部ストレージ180を指示せず、外部ストレージ180の接続指示のみをストレージ130に指示し、ストレージ130で全ポート141から検出した全外部ストレージ180各自が有する全デバイスを外部デバイスとして登録しても構わない。また、ストレージ管理者や管理サーバ110からは特に明示的な接続指示を与えず、ストレージ130に外部ストレージ180が接続されたのを契機に、ストレージ130が検出した全デバイスを外部デバイスとして登録してもよい。

【0177】

図9は、論理デバイス定義処理255の処理フローの一例を示す図である。論理デバイス定義処理255は、ストレージ管理者もしくは管理サーバ110からの指示を受けて、ストレージ130で定義される物理デバイス又は外部デバイス定義処理253で定義された外部デバイスに対して、上位又は下位論理デバイスを定義する処理である。

【0178】

ストレージ管理者もしくは管理サーバ110から論理デバイス定義指示を受け付けたSTは、当該指示をMAに送信する。論理デバイス定義指示には、論理デバイスが割り当てられる対象となる物理デバイス又は外部デバイスの番号、割り当てられる上位論理デバイスの番号、割り当てられる下位論理デバイスの番号及び当該下位論理デバイスを管理するCAの番号とが付加される。なお本実施形態では、説明の簡略化のため、1つの物理もしくは外部デバイスに対して1つの論理デバイスを割り当てるものとするが、2つ以上の物理もしくは外部デバイスからなるデバイスグループに対して1つの論理デバイスを、又は1つの物理もしくは外部デバイスに対して2つ以上の論理デバイスを、又は2つ以上の物理もしくは外部デバイスからなるデバイスグループに対して2つ以上の論理デバイスを定義しても構わない。

【0179】

ただし、それぞれの場合、下位論理デバイス管理情報201に、物理もしくは外部デバイス内での当該論理デバイスの開始位置およびサイズなどの情報の付加、対応する物理および外部デバイス番号エントリの複数デバイス化、物理および外部デバイス管理情報202、205に、対応する下位論理デバイス番号エントリの複数デバイス化を行うことがそれぞれ必要となる（ステップ901）。

【0180】

論理デバイス定義指示を受信したMAは、論理デバイス定義指示に付加されている情報より対象CAを特定し、特定されたCAに論理デバイス定義指示を送信する（ステップ902）。

【0181】

論理デバイス定義指示を受信したCAは、指示された物理もしくは外部デバイスに対して下位論理デバイスを登録する（ステップ903）。具体的にCAは、下位論理デバイス管理情報201の対象デバイスエントリ（論理デバイス定義指示に付加されている下位論理デバイスの番号に対応するエントリ）について、論理デバイス定義指示に付加されている下位論理デバイス番号をエントリ51に、論理デバイス定義指示に付加されている物理もしくは外部デバイスの番号をエントリ53に、下位論理デバイスのサイズをエントリ52に、論理デバイス定義指

示に付加されている上位論理デバイス番号をエントリ55に各々登録し、エントリ54には「オンライン」を設定する。

【0182】

また、当該物理／外部デバイスの対応CA番号／下位論理デバイス番号を設定し、デバイス状態を「オンライン」に更新する。登録が完了したらCAはMAにその旨通知する（ステップ903）。

【0183】

次にMAは、論理デバイス定義指示によって指定された下位論理デバイスに上位論理デバイスを割り当て、制御情報が更新された旨を他の各部位に通知する。具体的にMAは、上位論理デバイス管理情報203の当該デバイスエントリ（論理デバイス定義指示に付加されている上位論理デバイスの番号に対応するエントリ）について、論理デバイス定義指示に付加されている上位論理デバイスの番号をエントリ31に、上位論理デバイスのサイズをエントリ32に、論理デバイス定義指示に付加されているCA番号と下位論理デバイス番号をエントリ33に設定し、デバイス状態としてエントリ34には「オフライン」状態を設定し、エントリ35及び36に関しては未割当のため無効値を設定する。

【0184】

更に、当該上位論理デバイスの割当対象が外部デバイスの場合には、論理デバイス定義指示に付加されている外部デバイス番号をエントリ38に設定し、デバイスアクセスモードを初期値として「PA直結」に設定する。更に、当該上位論理デバイスの割り当て対象が外部デバイスである場合には、外部デバイス管理情報205の対応エントリ73に当該上位論理デバイス番号を設定する。そして、MAは制御情報の更新があった旨を各PA及びSTに通知する（ステップ904）。

【0185】

制御情報更新を通知されたPAは更新のあった情報をメモリ143へ取り込む。又、制御情報更新を通知されたSTは、更新のあった情報を取り込んだ後、論理デバイス定義処理の完了を要求元であるストレージ管理者若しくは管理サーバ110へ報告すべく終了報告を出力する（ステップ905）。

【0186】

図10は、LUパス定義処理252の処理フローの一例を示す図である。ストレージ管理者もしくは管理サーバ110からLUパス定義指示を受け付けたSTがMAに対して同指示を転送する。同指示には、対象上位論理デバイス番号、LUを定義するポート141の番号、LUNに加えて、同LUにアクセスするホスト100の識別情報（ホストが有するポート107のWWNなど）またはAAの識別情報（AA番号等）及び当該上位論理デバイスへの要求機能レベルを示す情報が付加される。

【0187】

ここで要求機能レベルとは、上位論理デバイスに要求されるデータの信頼性等のレベルを指す。具体的には、ストレージ130が提供するデータ複製やデバイス再配置などのデータ連携機能や、データをキャッシュ上に格納しておくキャッシュ常駐化によるアクセス高速化機能などの適用有無が指定される。これによつて、機能の適用の有無により、データの信頼性のレベルが変化する。なお、要求機能レベルとして、単に、当該上位論理デバイスに格納されるデータがCA経由で処理が必要か否か（たとえ、CAに接続される記憶装置157にデータが格納されなくても）を示す情報を指定してもよい（ステップ1001）。

【0188】

LUパス定義指示を受信したMAは、LUパス定義指示で指定される上位論理デバイスに関する上位論理デバイス管理情報203内エントリ38より、当該上位論理デバイスに対応するデバイスが外部デバイスか否かを判定し（ステップ1002）、外部デバイスの場合は外部デバイス接続変更処理を行う（ステップ1003）。

【0189】

外部デバイス接続変更処理完了後又はLUパス定義指示で指定される上位論理デバイスが外部デバイスに対応付けられていない場合、MAは、当該上位論理デバイスに対してLUパス登録を行う。具体的にMAは、上位論理デバイス管理情報203の当該デバイスエントリのエントリ35及びエントリ36にLUパス定義指示で指定される対応情報を設定し、LUパス管理情報206のLUパス定義

指示で指定されるポート141に関する空きエントリに、ターゲットID/LUNを始めとする構成情報を、LUパス定義指示に従って設定する。登録及び設定が完了したら、MAはその旨を他の各部位へ通知する（ステップ1004）。

【0190】

通知を受信したPAは、新たに設定、登録された情報を取込み、通知を受信したSTは新たに設定、登録された情報を取り込むと共に、要求元である管理サーバ110やストレージ管理者へ完了報告を通知すべく完了報告を出力する（ステップ1005）。

【0191】

図14は、外部デバイス接続変更処理256の処理フローの一例を示す図である。外部デバイスに対応する上位論理デバイスに対してLUパス定義される際に、当該外部デバイスに対してストレージ管理者が要求する機能レベルに基づきCA経由での入出力処理が必要であると判断された場合又はMAがSTを介してストレージ管理者もしくは管理サーバ110から上位論理デバイスに対する機能要求レベルの変更指示を受信した場合に、初期状態のPA直結接続から接続形態を切り替えるため、外部デバイス接続変更処理256が実行される。

【0192】

機能要求レベルの変更指示を受信したMAは、当該変更指示内の情報によって特定される上位論理デバイスの接続形態をCA経由とすべきか否かを、当該変更指示に基づいて判断する（ステップ1401）。CA経由とすべきである場合には、MAは、当該上位論理デバイスのアクセスモードがPA直結モードとなっているかどうかを上位論理デバイス管理情報203を参照して判断する（ステップ1402）。

【0193】

該当する上位論理デバイスがPA直結モードである場合には、MAは当該上位論理デバイスの接続形態を「PA直結」から「CA経由」へ変更する（ステップ1403）。既にCA経由モードとなっている場合には変更は必要ないので処理を終了する。

【0194】

ステップ1401で当該上位論理デバイスの接続形態をPA直結とすべき場合、MAは上位論理デバイス管理情報203を参照して当該上位論理デバイスのデバイスアクセスモードがCA経由となっているかPA直結となっているかを判断する（ステップ1404）。

【0195】

当該上位論理デバイスのデバスアクセスモードがCA経由となっている場合、MAは接続形態を「CA経由」から「PA直結」へ変更する必要がある。そこでまずMAは、当該上位論理デバイスの上位論理デバイス管理情報203内のエントリ37を「PA直結移行中」へ変更し（ステップ1405）、当該上位論理デバイスに対応付けられている下位論理デバイスを管理しているCAを上位論理デバイス管理情報203を参照して特定する。そして特定されたCAに当該上位論理デバイスの接続形態がCA経由からPA直結に変更される旨を通知する（ステップ1406）。

【0196】

通知を受けたCAは、対応する下位論理デバイスの全データについてディスクキャッシュ154を検索しディスクキャッシュをミス化、すなわち、物理デバイス又は外部デバイスへ未更新のデータをディスクキャッシュ154から対応デバイスへ書き出してデバイス内のデータを更新した後、更新済みのデータについては即、当該データに割り当てられたキャッシュ領域を解放する。CAは、ディスクキャッシュ154の検索をキャッシュ管理情報を用いて、当該下位論理デバイスの先頭から順次行う。又、CAは、検索およびミス化処理が終わった部分については、上位論理デバイス及び下位論理デバイス管理情報の切替進捗ポインタを進めることで進捗を管理する（ステップ1407）。

【0197】

ディスクキャッシュ154の全領域についてミス化処理が完了したら、CAはその旨をMAへ報告し（ステップ1408）、MAで上位論理デバイス管理情報203内のエントリ37を「PA直結」へ変更して接続形態変更処理を完了する。

【0198】

以上の実施形態によれば、ホスト100がコマンドを送付するアドレスが付与されるI/F部分（PA）とファイルアクセス処理を行うNASサーバ部分（AA）とデバイスの処理を行うキャッシュ制御部（CA）と外部ストレージ180と接続されるI/F部分（PA）を分離し、相互結合網170で接続することで、処理負荷を分散可能なNASシステムが実現できる。

【0199】

また、本実施形態によれば、ホスト100がコマンドを送付するアドレスが付与されるI/F部分（PA）とファイルアクセス処理を行うNASサーバ部分（AA）とデバイスの処理を行うキャッシュ制御部（CA）と外部ストレージ180と接続されるI/F部分（PA）とで「連携」モードを設け、大量のデータがホスト100とAAで送受されるような場合、対ホストのI/F部分（PA）とキャッシュ制御部（CA）もしくは外部ストレージ180と接続されるI/F部分（PA）間で直接データを送受することができ、AAの処理負荷を軽減できると同時に、ホストに対するコマンド応答が早いNASシステムが実現できる。

【0200】

また、本実施形態によれば、NAS管理情報223で、AAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係を、1つのPAに対し複数のAAが処理するようにエントリ1602に設定でき、ホスト100からは1台のNASシステムにアクセスするように見えながら、複数のAAで処理を実施することで処理負荷を軽減し、ホスト100に対するコマンド応答が早いNASシステムが実現できる。

【0201】

また、本実施形態によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係を、1つのAAが複数のIPアドレスを持っているようにエントリ1602に設定でき、ホスト100からは複数のNASシステムが存在するように見えながら、複数のPAまたはポート141でI/F処理を実施することで、処理負荷を軽減し、ホスト100に対するコマンド応答が早いNASシステムが実現できる。

【0202】

また、本実施形態によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係を、1つのAAが複数のIPアドレスを持っているようにエントリ1602に設定でき、ホスト100からは複数のNASシステムが存在するように見え、NASシステム毎に用途を使い分けることができるNASシステムが実現できる。

【0203】

また、本実施形態によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係の一部を他のAAに移すことができ、これにより、ホスト100に影響を及ぼすことなく、処理負荷が大きくなってきたAAを増設し処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0204】

また、本実施例によれば、NAS管理情報223でAAが管理するファイルとそれにアクセスするためにホスト100がコマンドを送信する先であるI/F部分のIPアドレスの対応関係の一部を変更することができ、これにより、処理負荷が大きくなってきたPAを増設もしくは処理負荷が少ないPAへ負荷を分散させ、これにより処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0205】

また本実施形態によれば、NAS管理情報223のAAに関するエントリ全てを新規のAAのエントリにコピーすることで、ホスト100に影響を及ぼすことなく、AAを処理能力の高い新しいパッケージに交換でき、処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0206】

また本実施形態によれば、ポート管理情報224のポートに付与されているIPアドレスの一部を別のポート141のエントリに移動することで、ホスト100

の設定を変更することなく、ポート141へのアクセスを分散させ、処理負荷を軽減し、ホストに対するコマンド応答を向上できるNASシステムが実現できる。

【0207】

また本実施形態によれば、AAに障害が発生した場合に、NAS管理情報223に予め設定してあった代替AAへ当該AAの管理情報を引き継ぐことができ、ホスト100の設定を変更することなく処理を継続可能なNASシステムが実現できる。

【0208】

また本実施形態によれば、PAのポート141に障害が発生した場合に、ポート管理情報224の当該ポート141に関する情報を全てを使用されていないポート141のエントリにコピーするか、使用されているポート141に当該ポートに関する情報の一部を移動することで、ホスト100の設定を変更することなく使用できるNASシステムが実現できる。

【0209】

また、本実施形態によれば、CAが外部ストレージを自デバイスとしてAAに提供できるので、ストレージ130にディスク装置157を搭載せずに、安い外部ストレージを利用することができ、前述の本実施形態の効果を持った低価格のNASシステムを実現することができる。

【0210】

また、本実施形態によれば、CAを介さずに、AAが直接外部ストレージを使用することができ、前述の本実施形態の効果を持った低価格のNASシステムを実現することができる。

【0211】

【発明の効果】

本発明によれば、NASの使用負荷が上がった場合にも、レスポンスが低下しないように、処理負荷が高い部分の処理を分散するよう制御可能である。

【0212】

また、本発明によれば、障害が発生した場合にも障害発生部分の情報を他に引き継ぐことができ、計算機から当該NASへのアクセスを継続もしくは計算機の設

定を変更しないで継続利用できる。

【図面の簡単な説明】

【図 1】

システムのハードウェア構成の一例を示す図である。

【図 2】

ストレージ 130 のソフトウェア構成の一例を示す図である。

【図 3】

上位論理デバイス管理情報の一例を示す図である。

【図 4】

LU パス管理情報の一例を示す図である。

【図 5】

下位論理デバイス管理情報の一例を示す図である。

【図 6】

物理デバイス管理情報の一例を示す図である。

【図 7】

外部デバイス管理情報の一例を示す図である。

【図 8】

外部デバイス定義処理の一例を示す図である。

【図 9】

論理デバイス定義処理の一例を示す図である。

【図 10】

LU パス定義処理の一例を示す図である。

【図 11】

要求振り分け処理の一例を示す図である。

【図 12】

コマンド処理の一例を示す図である。

【図 13】

ライトアフタ処理の一例を示す図である。

【図 14】

外部デバイス接続変更処理の一例を示す図である。

【図15】

管理端末のソフトウェア構成の一例を示す図である。

【図16】

NAS管理情報の一例を示す図である。

【図17】

ポート管理情報の一例を示す図である。

【図18】

NAS管理情報変更処理の一例を示す図である。

【図19】

ポート管理情報変更処理の一例を示す図である。

【図20】

NAS／ポート障害処理の一例を示す図である。

【図21】

ファイルアクセス処理の一例を示す図である。

【図22】

連携モード処理の一例を示す図である。

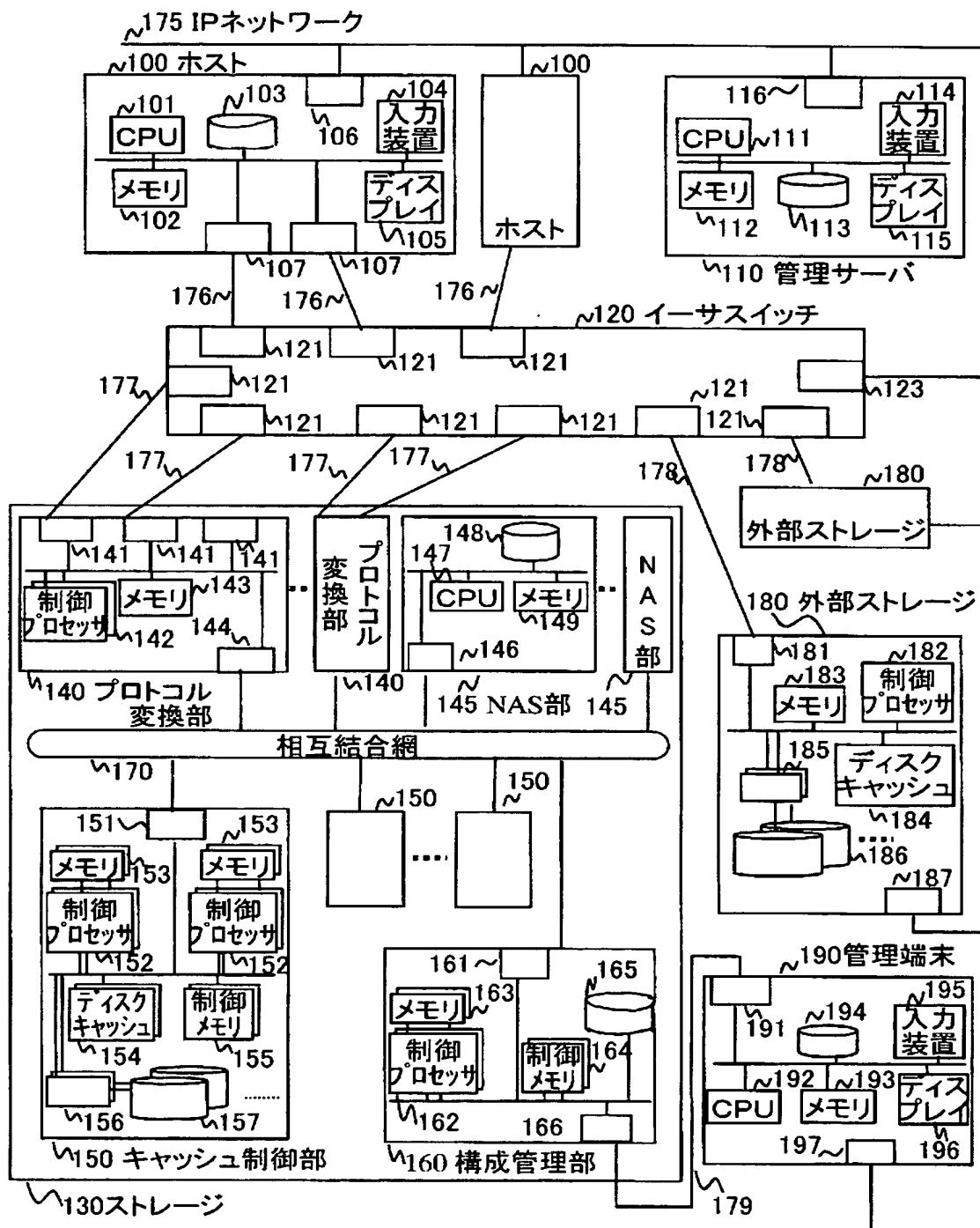
【符号の説明】

100…ホスト、110…管理サーバ、120…スイッチ、130…ストレージ装置、140…プロトコル変換部、145…NAS部、150…キャッシュ制御部、160…構成管理部、170…相互結合網、180…外部ストレージ装置、190…管理端末。

【書類名】 図面

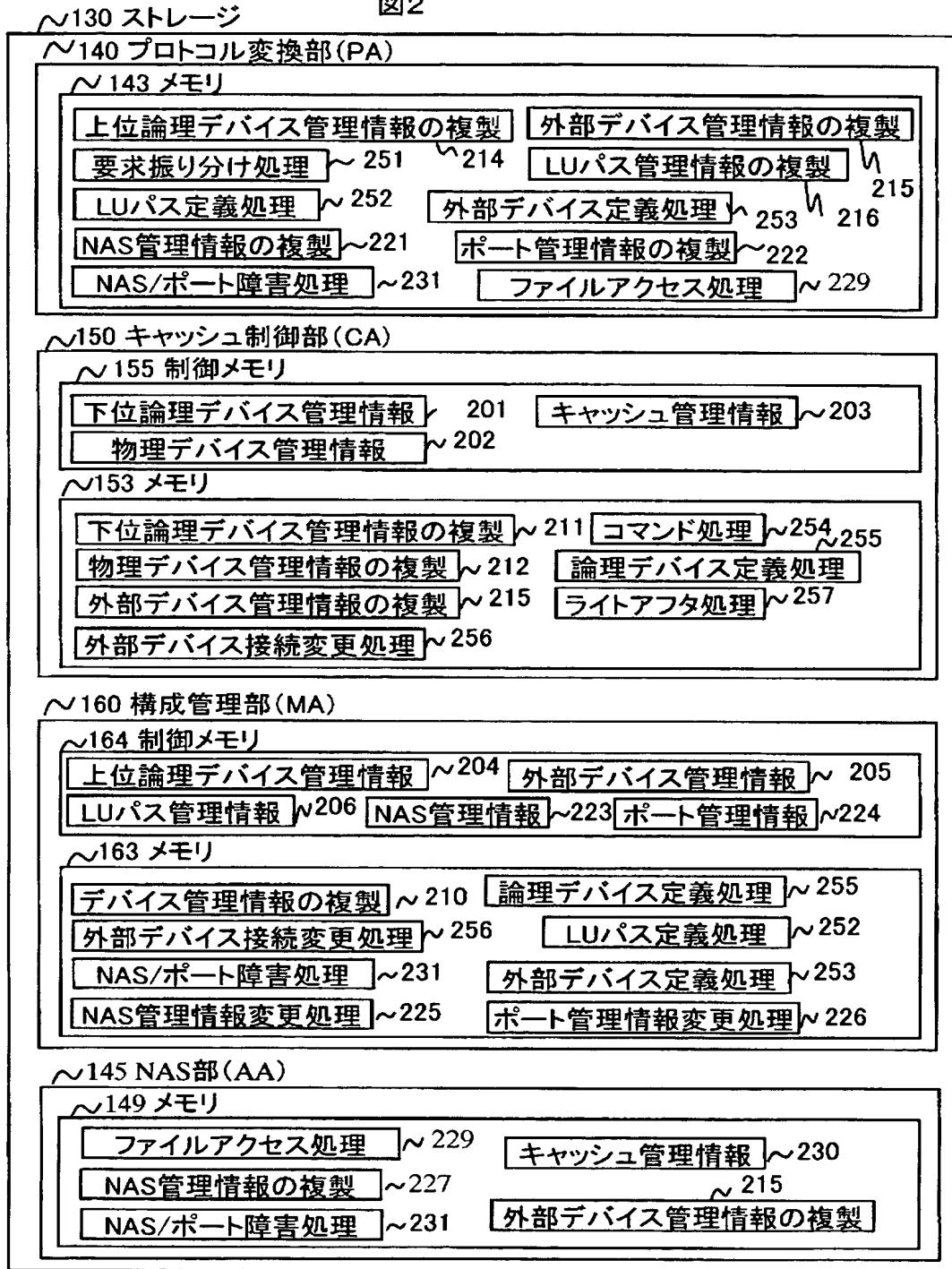
【図 1】

図1



【図2】

図2



【図3】

図3

上位論理デバイス管理情報204上位論理デ
バイス#0の
構成情報

上位論理デバイス番号	~31
サイズ	~32
対応CA番号、下位論理デバイス番号	~33
デバイス状態	~34
ポート番号、ターゲットID、LUN	~35
接続ホスト名	~36
デバイスアクセスモード	~37
対応外部デバイス番号	~38
切替進捗ポインタ	~39
：	

【図4】

図4

LUNパス管理情報206ポート/AA番号
#0の構成情報

ターゲットID/LUN	~41
対応上位論理デバイス番号	~42
接続ホスト名	~43
：	

【図5】

図5

下位論理デバイス管理情報201

下位論理デバイス#0の構成情報	下位論理デバイス番号	~51
	サイズ	~52
	対応物理/外部デバイス番号	~53
	デバイス状態	~54
	対応上位論理デバイス番号	~55
	切替進捗ポインタ	~56
	⋮	

【図6】

図6

物理デバイス管理情報202

物理デバイス#0の構成情報	物理デバイス番号	~61
	サイズ	~62
	対応下位論理デバイス番号	~63
	デバイス状態	~64
	RAID構成(RAIDレベル, データ/パリティ数)	~65
	ストライプサイズ	~66
	ディスク番号リスト	~67
	ディスク内開始オフセット	~68
	ディスク内サイズ	~69
⋮		

【図7】

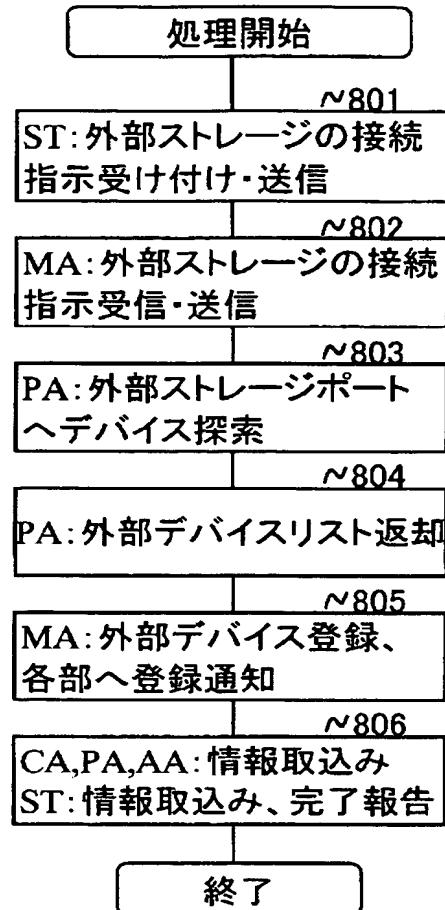
図7

外部デバイス管理情報205外部デバイス#0
の構成情報

外部デバイス番号	\\71
サイズ	\\72
対応上位論理デバイス番号	\\73
対応CA番号／下位論理デバイス番号／AA番号	\\74
デバイス状態	\\75
ストレージ識別情報	\\76
外部ストレージ内デバイス番号	\\77
PA番号／イニシエータポート番号リスト	\\78
ターゲットポートID／ターゲットID／LUNリスト	\\79
⋮	

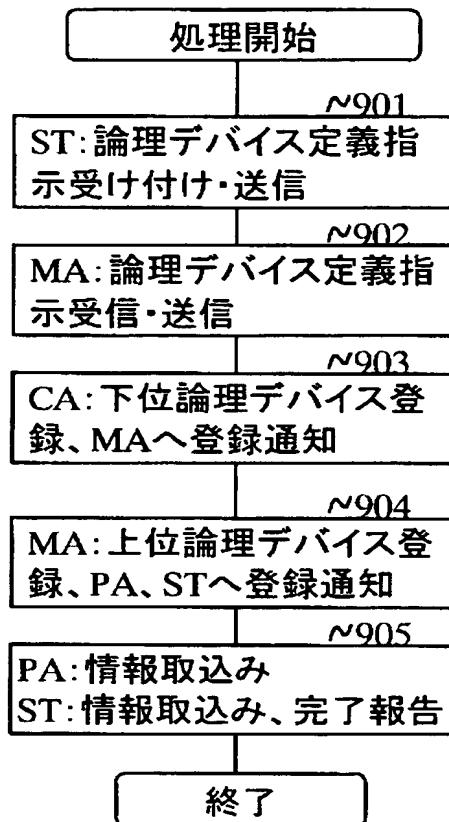
【図8】

図8

外部デバイス定義処理253

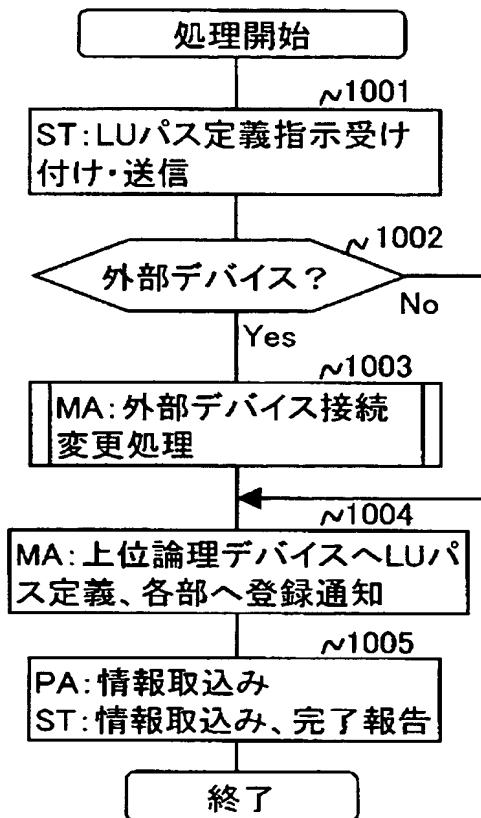
【図9】

図9

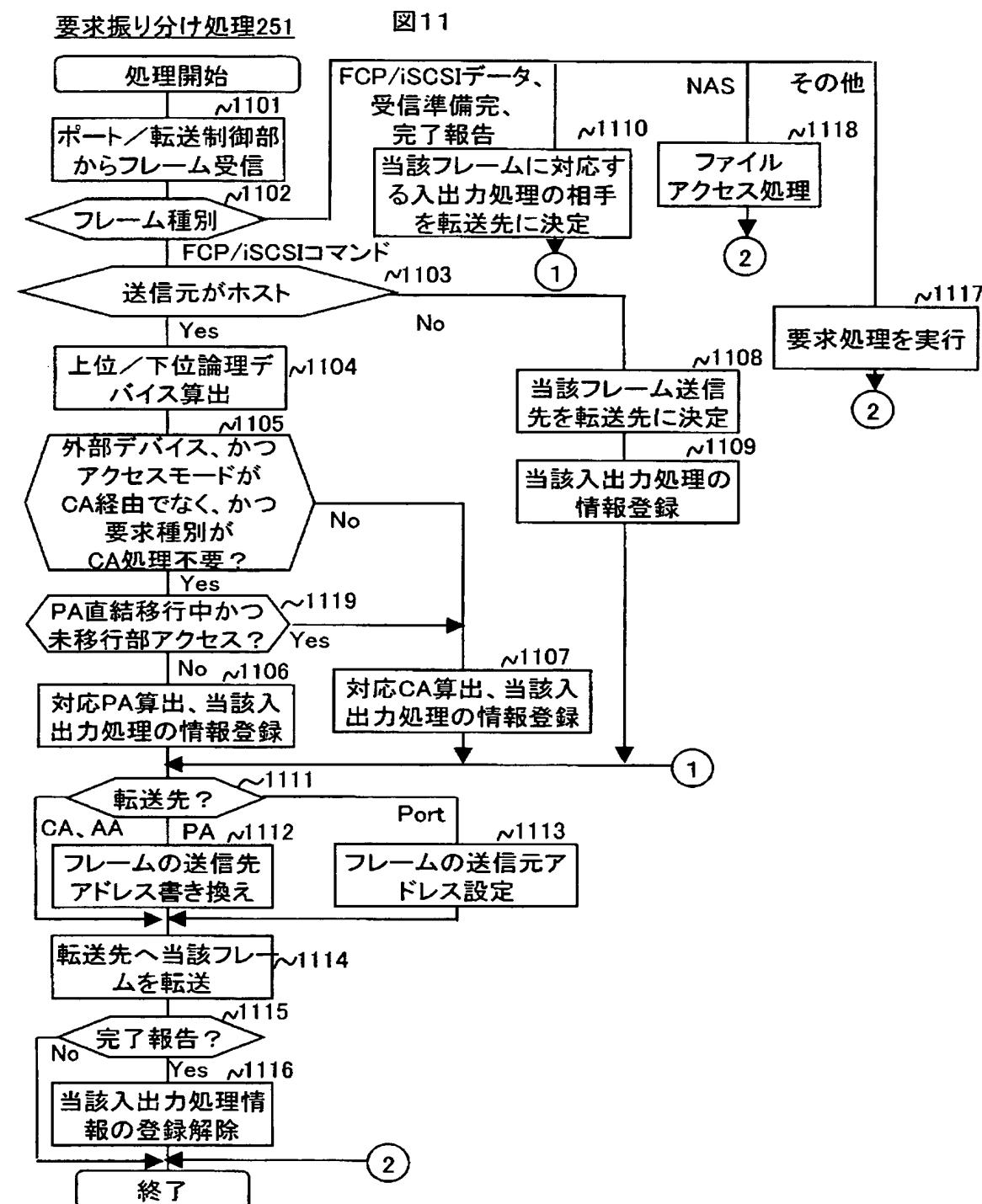
論理デバイス定義処理255

【図10】

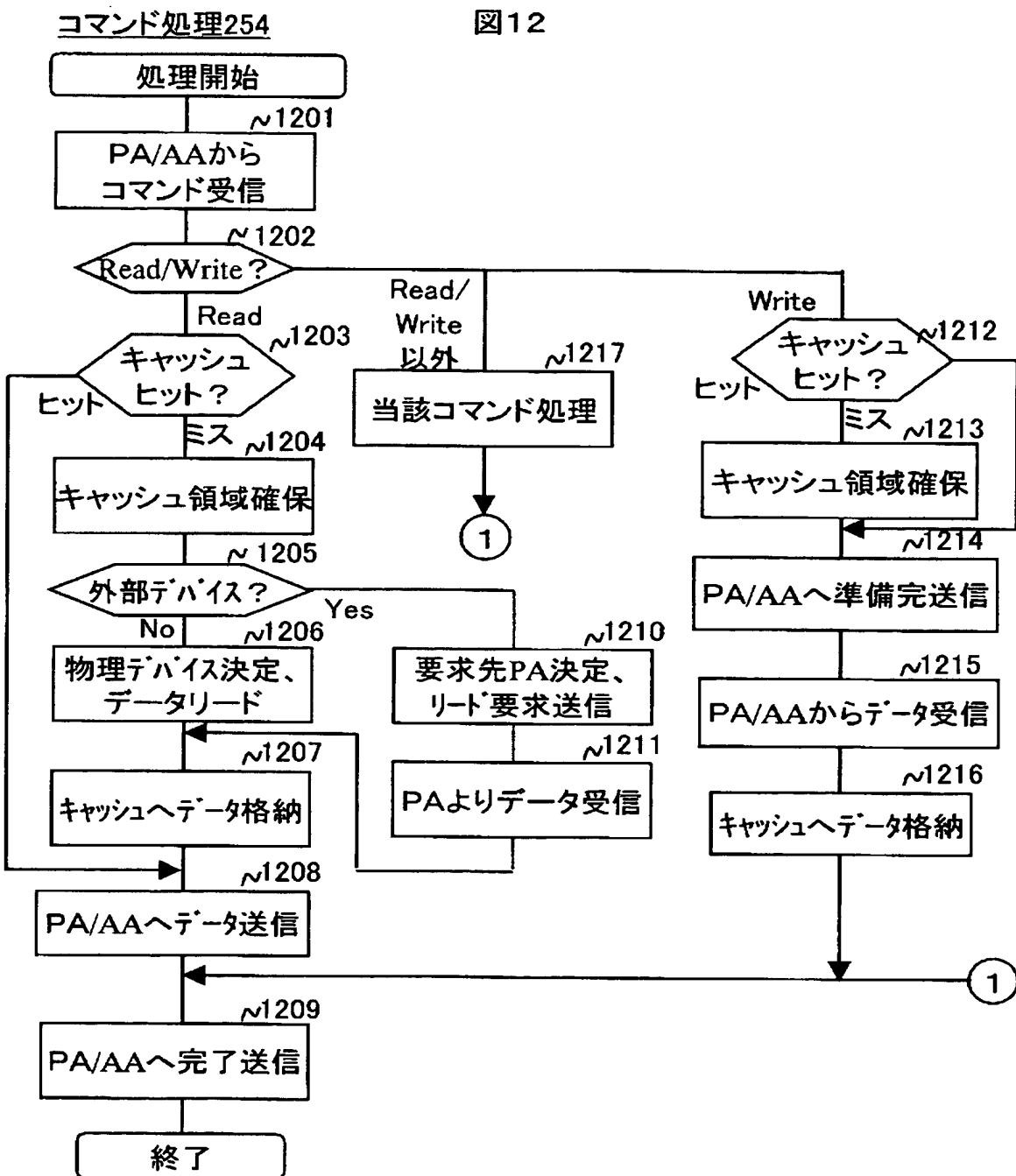
図10

LUパス定義処理252

【図11】



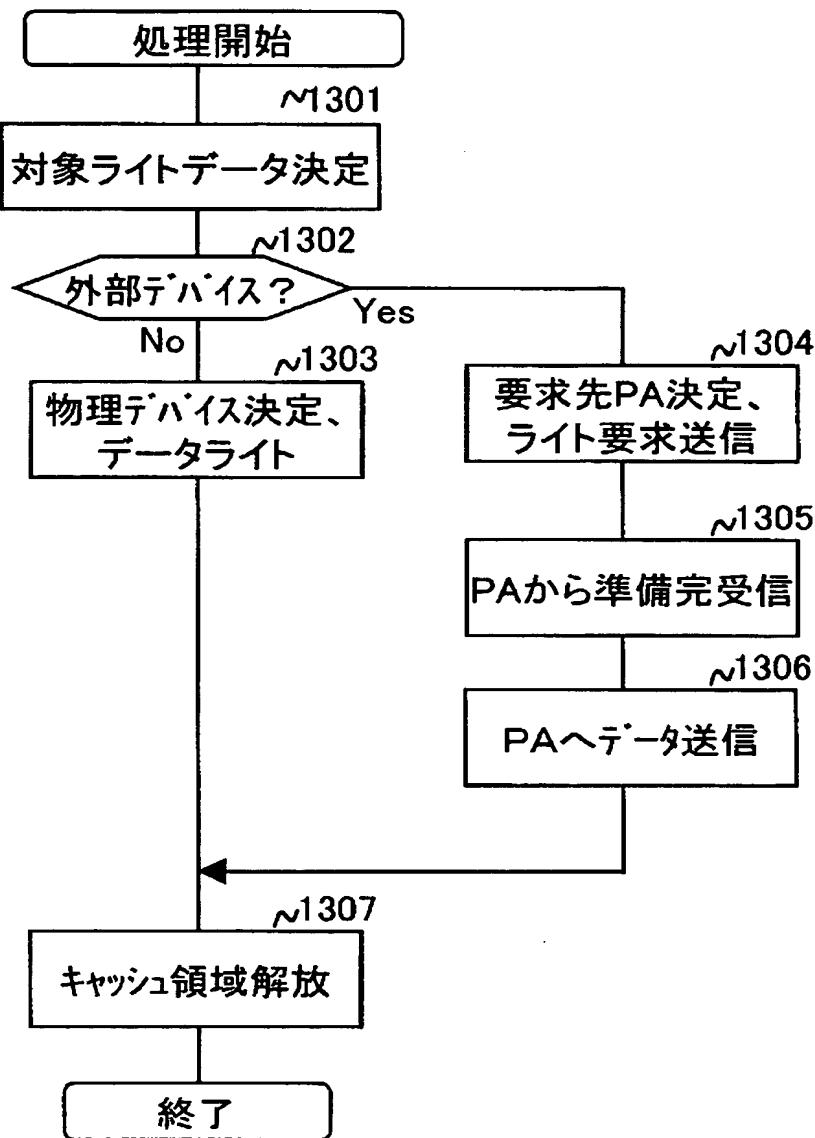
【図12】



【図13】

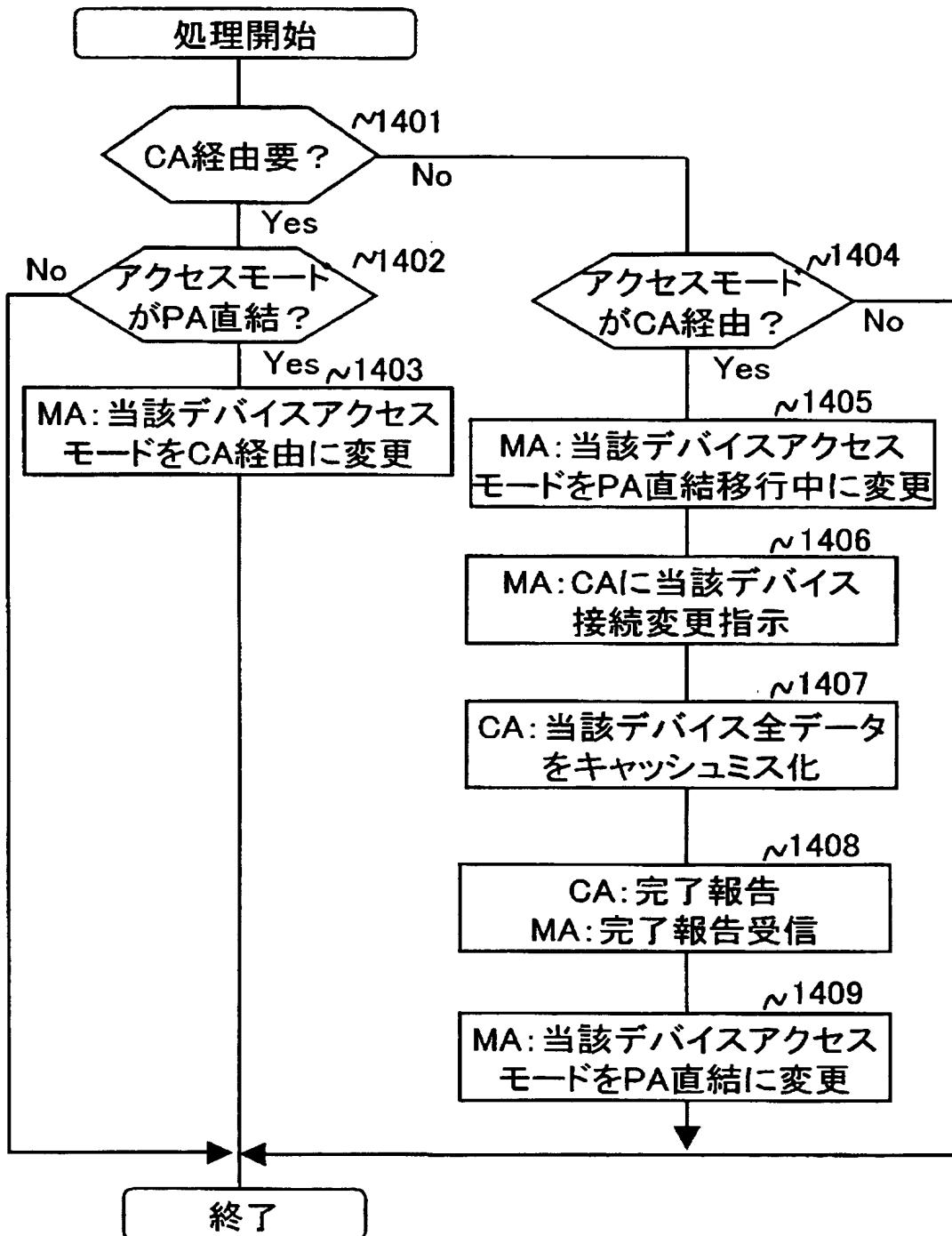
ライトアフタ処理257

図13



【図14】

図14

外部デバイス接続変更処理256

【図15】

図15

～190 管理端末(ST)
～193 メモリ
デバイス管理情報の複製 ～210
NAS管理情報の複製 ～231
論理デバイス定義処理 ～255
NAS管理情報変更処理 ～225
NAS/ポート障害処理 ～231
LUパス定義処理 ～252
ポート管理情報の複製 ～32
外部デバイス定義処理 ～253
ポート管理情報変更処理 ～226

【図16】

NAS管理情報 223

図16

AA#0の構成情報

AA番号	～1601
IPアドレス、ディレクトリ名、対応CA番号、上位論理デバイス番号／外部デバイス番号のリスト	～1602
状態	～1603
アクセスモード	～1604
代替AA番号	～1606
⋮	

【図17】

ポート管理情報 224

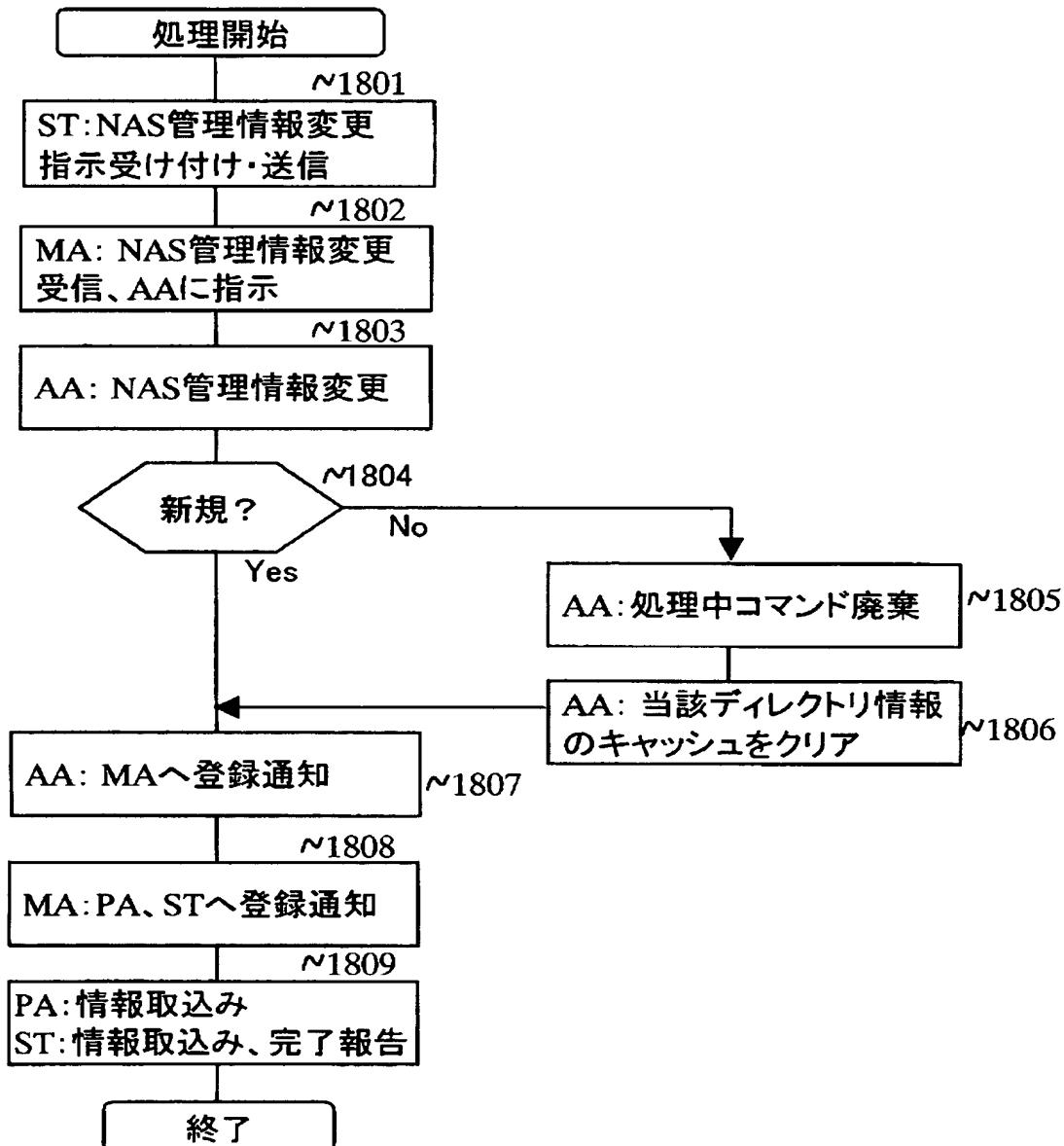
図17

ポート番号 #0
の構成情報

ポート番号	~1701
MACアドレス	~1702
IPアドレスのリスト	~1703
プロトコル	~1704
状態	~1705
アクセスモード	~1706
対応外部デバイス番号	~1707
代替ポート番号	~1708
⋮	

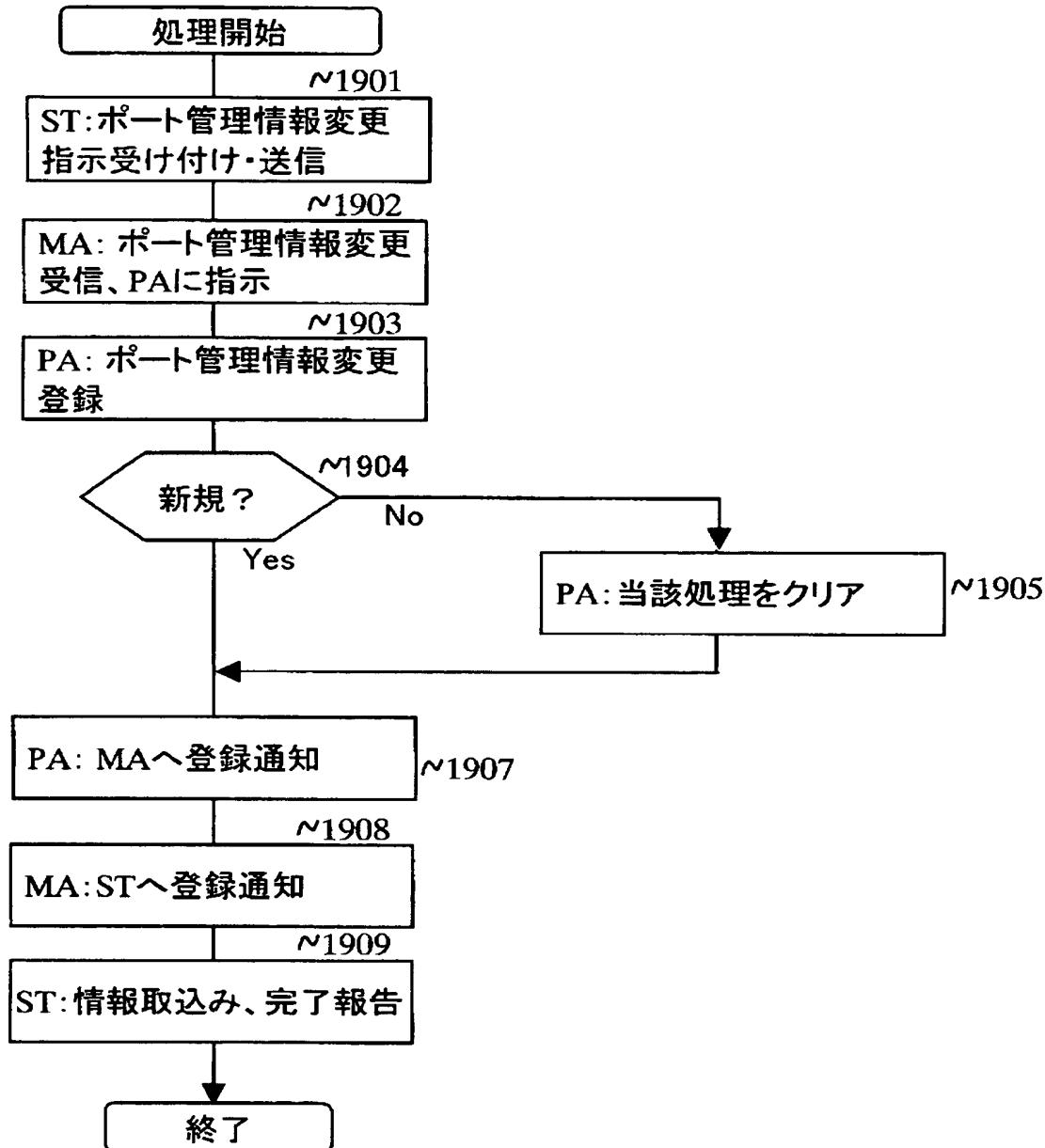
【図18】

図18

NAS管理情報変更処理225

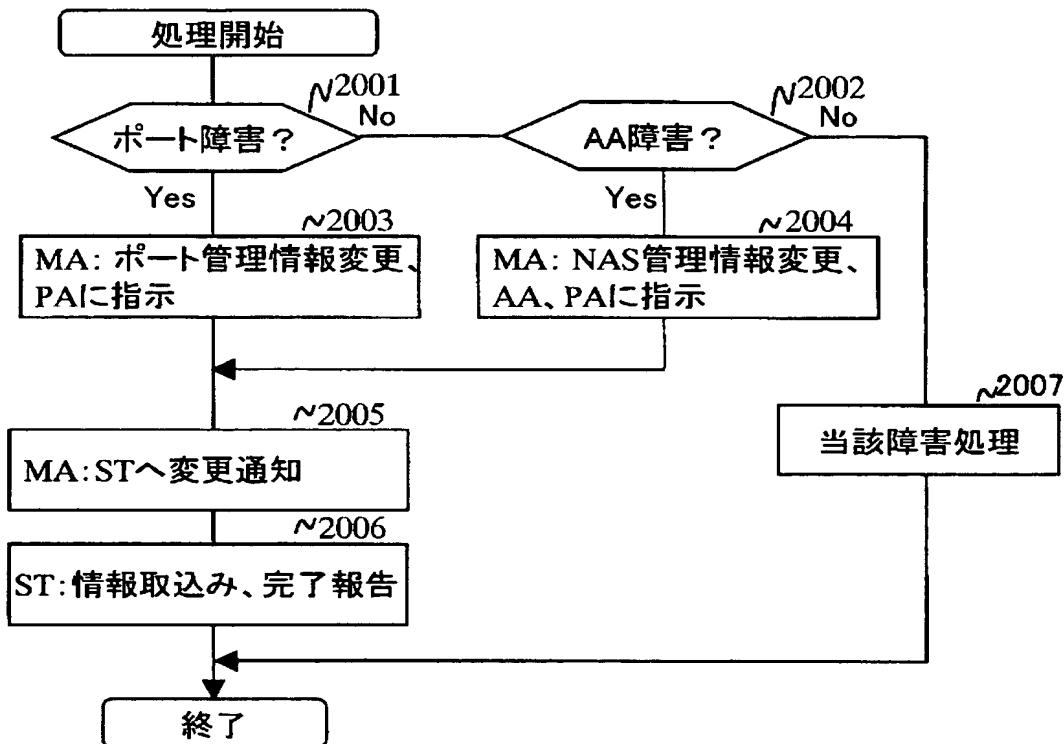
【図19】

図19

ポート管理情報変更処理226

【図20】

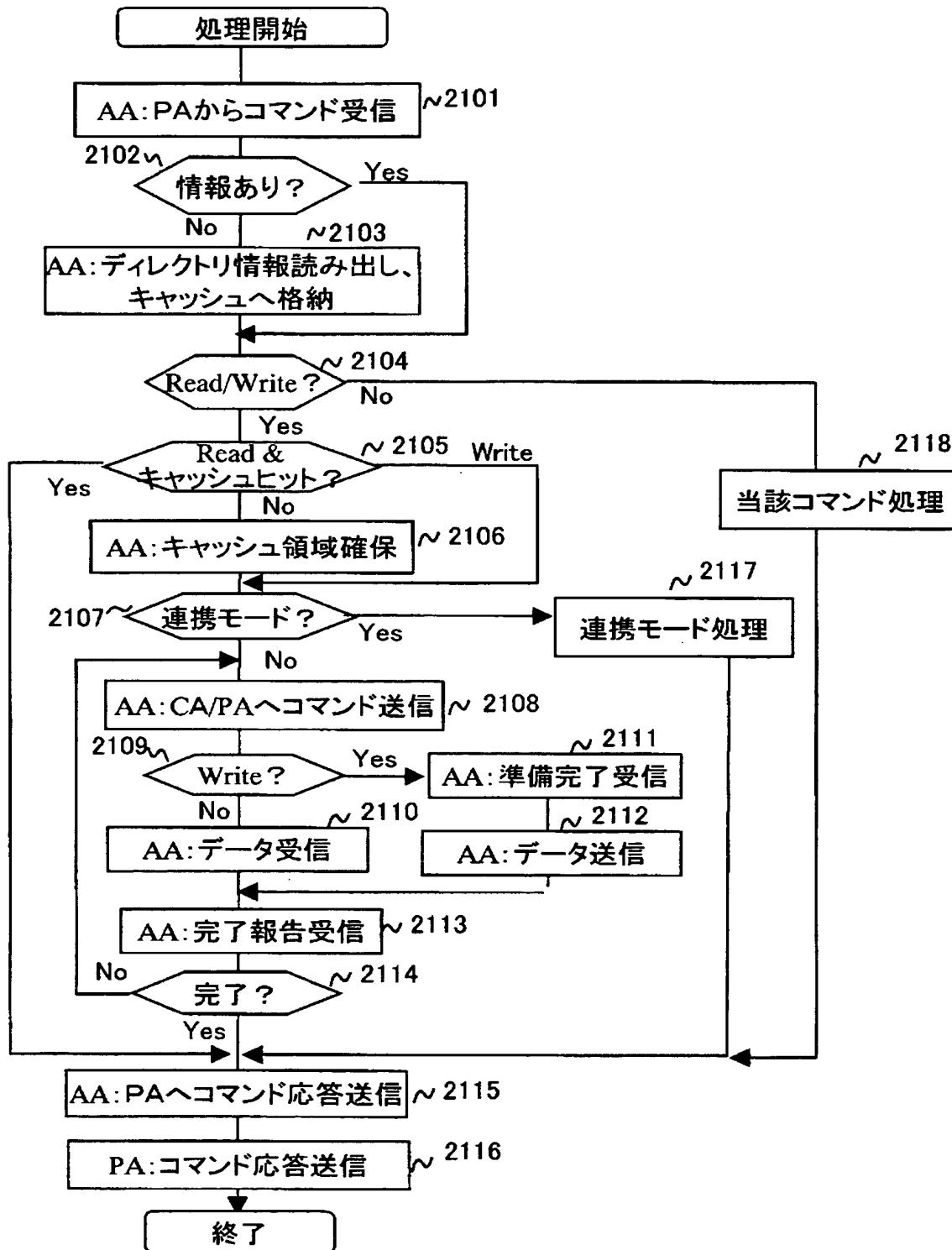
図20

NAS/ポート障害処理231

【図21】

ファイルアクセス処理229

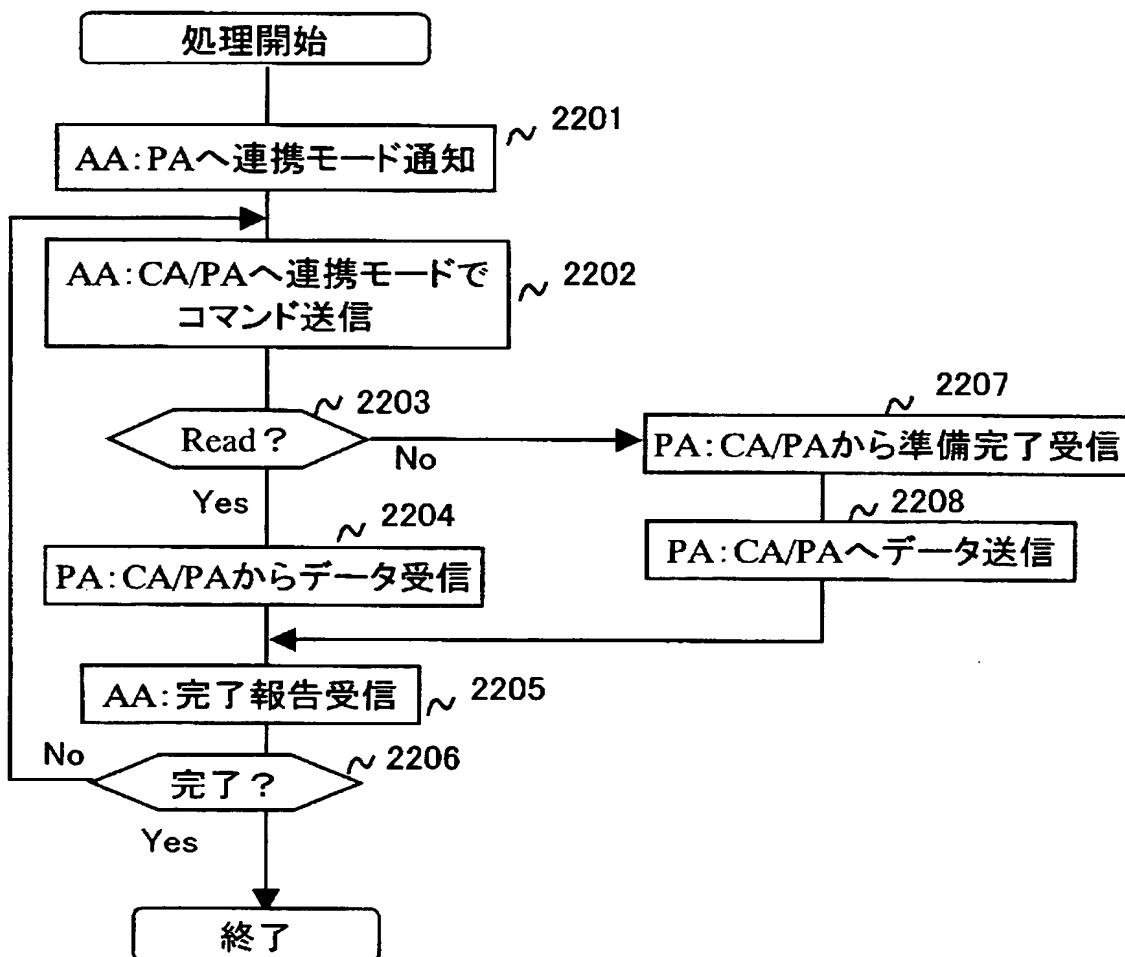
図21



【図22】

連携モード処理2118

図22



【書類名】要約書**【要約】****【課題】**

NASサーバと当該NASサーバに付与されたネットワークアドレスが設定されるポートとの関係が1：1で一意の決定されるためTCP/IP処理部またはNASサーバの処理性能がネックでレスポンスが低下する。

【解決手段】

ホストがコマンドを送付するアドレスが付与されるI/F部分とNASサーバ部とデバイスの処理を行うキャッシュ制御部と外部ストレージを接続するI/F部分を分離し、相互結合網170で接続することで、I/FとNASサーバの対応関係およびI/Fに付与されるアドレスとNASサーバとの対応関係を処理負荷等に応じ任意に変更する。

【選択図】 図1

認定・付加情報

特許出願の番号	特願2003-206168
受付番号	50301300042
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年 8月 7日

<認定情報・付加情報>

【提出日】	平成15年 8月 6日
-------	-------------

特願 2003-206168

出願人履歴情報

識別番号

[000005108]

1. 変更年月日

1990年 8月31日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台4丁目6番地

氏 名

株式会社日立製作所